



(12) EUROPEAN PATENT APPLICATION

②① Application number : 92303067.0

⑤ Int. Cl.⁵: **C12N 15/62**, C07K 3/18,
C07K 13/00, // C12N5/00

②② Date of filing : 07.04.92

(30) Priority : 19.04.91 US 687819

④3 Date of publication of application :
04.11.92 Bulletin 92/45

(84) Designated Contracting States :
AT BE CH DE DK ES FR GB GR IT LI LU MC NL
PT SE

71) Applicant : ROHM AND HAAS COMPANY
Independence Mall West
Philadelphia Pennsylvania 19105 (US)

(72) Inventor : **Cress, Dean Ervin**
104 Leslie Lane
Souderton, Pennsylvania 18964 (US)
 Inventor : **Haase, Ferdinand Carl**
53 Skyline Drive
Chalfont, Pennsylvania 18914 (US)

(74) Representative : Harding, Charles Thomas et al
Rohm and Haas (UK) Ltd., European
Operations Patent Dept., Lennig House, 2
Masons Avenue
Croydon CR9 3NB (GB)

(54) Hybrid polypeptide containing an avidin binding polypeptide.

(57) A hybrid polypeptide is disclosed comprising an avidin-binding polypeptide containing a biotin binding domain fused to a polypeptide of interest wherein the avidin-binding polypeptide is "upstream" of the polypeptide of interest. The hybrid polypeptide is produced by recombinant DNA techniques. The hybrid polypeptide may also contain a cleavage site for cleaving the polypeptide of interest from the avidin-binding polypeptide by using an appropriate proteolytic or chemical reagent. The hybrid polypeptide is expressed in appropriate host cells transformed with the DNA expression vector encoding the hybrid polypeptide, and may be recovered from crude cell extracts in high yield and high purity using avidin affinity chromatography. Following avidin affinity purification, the polypeptide for attachment and polypeptide of interest may be cleaved to yield polypeptide of interest in a highly pure and highly active state.

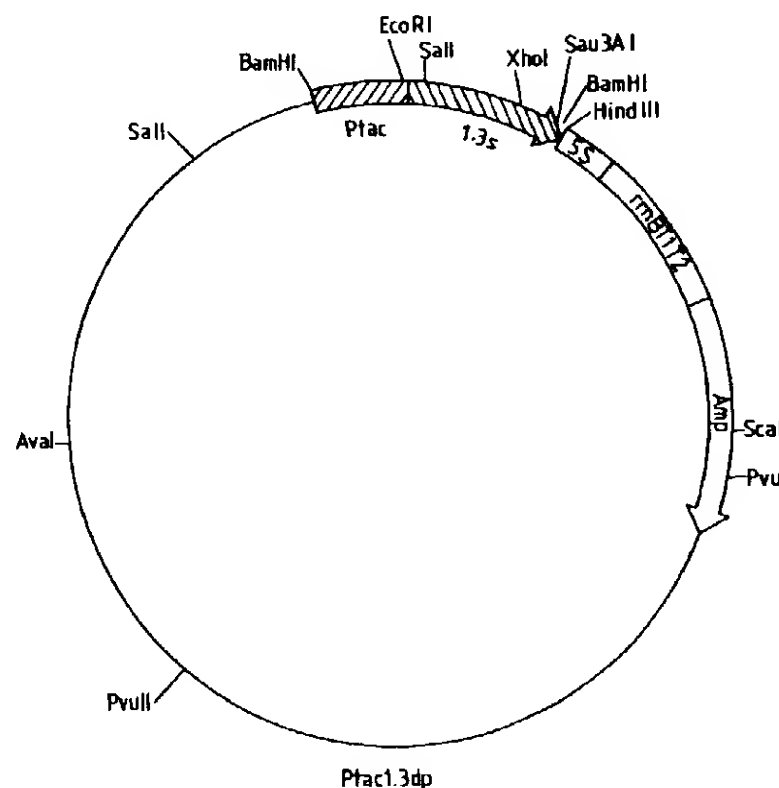


Fig.1.

The present invention relates to a hybrid polypeptide.

In particular, the present invention relates to a recombinant hybrid polypeptide comprising a polypeptide of interest fused to an avidin-binding polypeptide. The avidin-binding polypeptide contains a biotin attachment domain.

5 More particularly, the present invention relates to a hybrid polypeptide comprising a polypeptide of interest fused to a biotinylated polypeptide that can bind to avidin.

The present invention also relates to a nucleic acid sequence that encodes for the hybrid polypeptide, a process for producing the same and a process for recovering same.

In particular, the nucleic acid sequence is a DNA expression vector.

10 Generally, the synthesis of commercially important peptides and proteins has been limited by high production and purification costs and, also, poor product recovery. Until recently, animals, micro-organisms, plants, cadavers, serum, and urine have been the only sources from which bioactive polypeptides could be purified.

However, advances in recombinant DNA technology have made the biological synthesis of valuable polypeptides possible and in commercial quantities. In this regard, recombinant DNA molecules directing the synthesis of commercially useful polypeptides can be introduced into procaryotic or eucaryotic expression systems. For example, recombinant DNA technology has enabled human growth hormone production by recombinant bacteria and, today, fermentation replaces the traditional source.

20 Todate, biological synthesis is the only practical approach to the commercial-scale synthesis of peptides of greater than 20 amino acid residues. Once synthesized, the desired polypeptide product must be purified from a complex mixture of cellular components. The degree of purification depends upon the intended application of the polypeptide. The cost of purification can account for up to 70% of the cost of production, as substantial losses of active ingredient usually occur during multistep purification processes.

Polypeptide purifications are usually achieved through one or more processes which are based upon physical properties of the polypeptide of interest. For example, proteins may be separated on the basis of solubility, size, ionic properties or affinity for specific ligands; usually several of these techniques are required to achieve acceptable purity.

Affinity resin chromatography can greatly reduce the number of purification steps required to achieve the desired level of purity. Affinity purification is based upon a specific binding interaction between a polypeptide to be purified and a ligand which is usually attached to a solid support. As used herein, the polypeptide binds 30 to the ligand by virtue of a prosthetic group bound to an attachment domain present on the polypeptide. When a complex mixture such as a cell extract or crude mixture of synthetic peptides is passed over an affinity resin, the polypeptide to be purified is selectively retained by the resin and all molecules lacking the prosthetic group on the attachment domain are washed away from the resin.

Therefore, in a single step, the polypeptide of interest may be recovered in high purity.

35 In order to use affinity chromatography to advantage for polypeptide purification, recombinant DNA technology can be used to construct chimeric gene fusions for recombinant hybrid polypeptides which in bacterial host cells incorporate the following elements: a 5' promoter; DNA coding for a polypeptide of interest; DNA coding for a polypeptide that contains a ligand binding domain; and optionally ribosomal terminators, such as the rrnB terminators found on the E. coli expression vector pkk223-3 (Brosius, j. and Holy A., Proc Nat Acad Sci USA 81:6929-6933 (1984); Brosius, j. et al Plasmid 6:112-118 (1984)).

Suitable promoters are those which maximize expression of the desired gene in the host cell, and factors to be considered in promoter construction are discussed by Old and Primrose in Chapter 7 of Principles of Gene Manipulation 3rd Edition (Blackwell Scientific Publications, Palo Alto CA 1985). Examples of bacterial promoters appropriate for expression of cloned genes include the PL, tac, lac, and trp promoters (ibid).

45 The DNA used to construct chimeric gene fusions can be obtained from organisms or can be novel synthetic DNA fragments, or combinations thereof. The DNA sequences are assembled into a chimeric gene, which is inserted into a DNA expression vector in such a manner that in the appropriate host organism, the polypeptide of interest and the polypeptide for attachment to the affinity resin are produced as a single polypeptide chain.

Other systems for affinity purification of hybrid recombinant polypeptides are known. However, significant technical obstacles limit their use for commercial-scale polypeptide purification. For example, chimeric genes encoding polypeptides containing a polyarginine C-terminal tail (Sassenfeld and Brewer, Biotechnology 2:76-81 (1984)) or polyhistidine domain (Smith et al. J. Biol Chem 263:7211 (1988)) can facilitate separation by ion exchange or metal chelate ion chromatography. Such systems are not broadly applicable because affinity interaction depends upon physical properties of the fusion polypeptide (chargeability to chelate metals), and it is not always possible to achieve sufficient change in these physical properties to permit affinity binding.

55 Another type of affinity chromatography is immunoaffinity chromatography, wherein polypeptides of interest are fused to immunogenic proteins such as E. coli beta-galactosidase (Ruther and Muller-Hill, EMBO J 2:1791-1794 (1983)) or small hydrophilic peptides (Hopp et al., US 4,703,004 (1988)) to achieve purification. Polypep-

tides fused to staphylococcal protein A can be purified using IgG-Sepharose (Nilsson et al. EMBO J. 4:1075-1080 (1985), Lowenadler et al. EMBO J. 5:2393-2398 (1986)). Polypeptides fused to Protein G can be isolated using albumin as the immobilized ligand (Nygren et al. J. Mol. Recognition, 1:69 (1988)).

A critical disadvantage limiting the usefulness of these prior art methods is that extreme conditions, including the use of denaturants, are necessary to remove the fusion proteins from the affinity resin, which may destroy biological activity if native folding cannot be achieved. Low product recovery rates can also limit the usefulness of such systems.

Affinity based upon the binding of small molecules by a large protein is known as substrate-affinity chromatography. A small molecule, a ligand, forms a complex with a specific ligate. Examples of ligand:ligate combinations include avidin:biotin (Green in Advances in Protein Chemistry Vol 29 pp 85-133 Anson et al., Eds.(1975)), streptavidin:biotin (PCT/US85/01901, Meade and Garvin (1985)), lipoic acid:avidin (Green *ibid*), chloramphenicol acetyl transferase:acetyl CoA (EPO 0131363, Bennet et al. (1984)), beta-galactosidase: para-aminophenyl-beta-D-thio-galactoside (Offensberger et al. Proc. Natl Acad. Sci USA 82:7540-7544 (1985)), phosphate binding protein:hydroxyapatite (Anba et al. Gene 53:219 (1987)), maltose binding protein:starch (EPO 286239, Guan et al. 1988), and glutathione S-transferase:glutathione (Smith and Johnson Gene 67:21-30 (1988)).

In recent years, the unique properties of the prosthetic group biotin and its exceptionally high affinity (10¹⁵ M⁻¹) and specificity for the proteins avidin and streptavidin (Green *ibid*.) have been exploited to devise powerful and widely applicable tools for microbiology, biochemistry and medical science (Wilchek and Bayer Analyt Biochem 171:1-32 (1988), Bayer and Wilchek Methods in Biochem Anal 26:1-45 (1980)).

Biotin is a prosthetic group found on only a few protein species (Ann N.Y. Acad. Sci 447:1-441, Dakshinamurti and Bhagavan, Eds. (1985)). Attachment *in vivo* is mediated by biotin holoenzyme synthetases which recognizes a highly conserved attachment domain and catalyzes the covalent attachment of biotin to that domain (Wood et al, J Biol Chem 225:7397-7409 (1980); Shenoy and Wood, FASCB S 2:2396-2401 (1988)).

Experiments using recombinant DNA technology have shown that biotin holoenzyme synthetases will biotinylate heterologous polypeptides containing this conserved attachment domain. For example, the 1.3S subunit of the enzyme transcarboxylase from *Propionibacterium*, which contains the conserved sequence, when cloned and expressed in *E. coli* is biotinylated by the *E. coli* synthetase (Murtif et al. Proc Nat Acad Sci USA 82:5617-5621 (1985)).

A polypeptide or part of a polypeptide containing the conserved biotin attachment domain, such as entire 1.3S (SEQ ID NO:1) protein or the biotin-binding recognition sequence identified within the 1.3S protein from *Propionibacterium*, (SEQ ID NO:2) can be incorporated into a hybrid recombinant polypeptide. Such a hybrid polypeptide containing a biotin attachment domain fused to one or more polypeptides of interest could be used to achieve the separation of virtually any recombinant protein based upon the affinity of the ligand avidin for the ligate biotin.

Avidin:biotin chromatography shares advantages generally applicable to substrate affinity chromatography systems for commercial-scale polypeptide purification. Substrate-affinity resins are generally inexpensive. Fusion proteins can be recovered using mild conditions by elution with free ligand.

Post-translational addition of the biotin prosthetic group is independent of the final folded state of the protein (Wood et al. J Biol Chem 255:7397-7409 (1980)), an advantage when the host cell performs no post-translational modifications on the recombinant polypeptide.

A ligand domain such as the domain directing biotin attachment would therefore be particularly advantageous for recovery of fusion proteins found in inclusion bodies or for recovery of insoluble proteins which require denaturants or zwitterionic detergents for solubilization during extraction, prior to affinity chromatography.

PCT WO 90/14431, which names Cronan as an inventor, discloses a hybrid DNA sequence encoding a fusion protein comprising a first DNA sequence which encodes an amino acid sequence that allows for post-translation modification of the fusion protein; and a second DNA sequence joined end to end with the first DNA sequence and in the same reading frame, the second DNA sequence encoding a selected protein or polypeptide. In each of the examples the first DNA sequence is fused to the 3' end of the second DNA sequence (i.e. the first DNA sequence is downstream of the second DNA sequence).

Also disclosed in PCT WO 90 / 14431 is a vector comprising the hybrid DNA, a host transformed with the vector, a method of producing a fusion protein by culturing the transformed host under conditions permitting expression of the fusion protein, a fusion protein comprising a selected protein or polypeptide linked to an amino acid sequence that allows for post-translation modification of the fusion protein and a method of isolating the fusion protein comprising providing a binding partner that binds to the fusion protein only after it has been modified, contacting the modified fusion protein with the binding partner under conditions permitting binding, separating the modified fusion protein bound to the binding partner from unbound materials in the mixture, and eluting the modified fusion protein.

The work of Cronan is also disclosed in J Biol Chem 265:10327-10333 (1990) (Cronan) wherein a recombinant DNA plasmid from E. coli (Murtif et al. Proc Nat Acad Sci USA 82:5617-5621 (1985)) was used to construct fusion genes containing segments of the 1.3S gene, which contain the biotin attachment domain.

5 Cronan (ibid) demonstrated that 1.3S sequences can be used to specifically label proteins in vivo, and to purify proteins from crude cell lysates by avidin affinity chromatography.

As in each of the examples of PCT WO 90/14431, Cronan's (ibid) chimeric genes were constructed by fusing the 3' end of the genes of interest to the 5' end of the 1.3S gene (i.e. the 1.3S gene is downstream of the gene of interest), yielding hybrid recombinant polypeptides having the polypeptides of interest fused to the N-terminus of the 1.3S polypeptide.

10 The PCT WO 90/14431 and Cronan fusions are consistent with the teachings of Murtif and Samols (J Biol Chem 262:11813-11816 (1987)) who teach the fusion of the 3' end of the gene of interest to the 5' end of the 1.3S gene (the N-terminus of the 1.3S polypeptide) to avoid interfering with the attachment of biotin to its binding domain. Murtif and Samols (ibid) teach that the conformation of the COOH terminus of the 1.3 S polypeptide, and the spatial relationship between this region and a lysine residue positioned exactly 35 residues from the COOH terminus position to which biotin is attached in vivo, are essential for proper enzymatic recognition and biotinylation of the 1.3S polypeptide.

Murtif and Samols (ibid) further teach that the conformation of the carboxyl terminal region of the 1.3S polypeptide is critical for biotinylation, and that altering the hydrophobicity of the carboxyl terminal region of the 1.3S polypeptide "eliminates biotinylation."

20 Murtif Samols did observe biotinylation of 1.3S polypeptides, each lengthened by two amino acids at the 1.3S carboxyl terminus. However, such additions of two amino acids to the C-terminus did not substantially change its hydrophobicity and such small additions would not be expected to change the conformation of the C-terminus.

US-A-4782137 discloses a series of recombinant DNA techniques for preparing a hybrid polypeptide consisting of an identification peptide and a desired functional protein. The identification peptide has an antigenic terminal and a cleavable linking portion disposed between the antigenic terminal portion and the protein molecule. The linked linking portion of the identification peptide is cleavable at a specific amino acid residue adjacent the functional protein by use of a sequence specific proteolytic enzyme or chemical agent. When the protein is cleaved from the isolated hybrid polypeptide, mature functional protein in a highly purified and active state is released. As with Murtif and Samols (ibid), PCT WO 90/14431 and Cronan (ibid), US-A-4782137 only discloses the use of a desired functional protein "upstream" of an identification protein.

US-A-4839293 discloses a process for preparing a fused gene consisting of a streptavidin gene fused to a gene encoding the human LDL receptor. Methods are also disclosed that utilise the fused gene to produce labelled, chemically modified proteins *in vivo* and also to isolate a protein knowing only the nucleotide sequence of the gene encoding the protein. The fused gene comprises a first DNA fragment encoding a target protein of interest fused to a second DNA fragment encoding streptavidin which has a multiplicity of binding sites for biotin or a biotin derivative. The fused gene of US-A-4839293 is capable of expressing a fused protein *in vivo* when the gene is inserted into a suitable expression vector and introduced into a suitable host cell. The fusion proteins are separated by the addition of biotin. Apparently, this method overcomes the disadvantages associated with the then-known commercial preparations of streptavidin as it utilises a biotin contaminant-free source of streptavidin which has all four valencies free for biotin binding.

However, as with Murtif and Samols (ibid), PCT WO 90/14431, Cronan (ibid) and US-A-4782137, US-A-4839293 only discloses the use of a desired functional protein "upstream" of an identification protein.

45 The problems associated with the prior art methods can therefore be summarised as follows. First, they do not allow proteins and the like to be isolated in a high level of purity. Second, they do not allow proteins and the like to be isolated in high yields. Third, they do not allow proteins and the like to be separated in high purity and yield in a single chromatographic step. Fourthly, they do not provide for the alteration of the carboxyl terminal region biotin-binding polypeptide part of the hybrid polypeptides. In fact, and as discussed above, the prior art teachings (such as those of Murtif and Samols (ibid), PCT WO 90/14431 and Cronan (ibid)) clearly suggest that altering, for example, the hydrophobicity of the carboxyl terminal region of the 1.3S polypeptide would eliminate biotinylation and therefore prevent the isolation of the hybrid polypeptide.

55 The solution to these problems, which the present invention provides, is a recombinant hybrid polypeptide (including methods and nucleic acid sequences for producing same) comprising a polypeptide of interest fused to an avidin-binding polypeptide containing a biotin attachment domain wherein the polypeptide of interest is fused to the C terminus of the avidin-binding polypeptide (i.e. the polypeptide of interest (i.e the desired functional protein) is "downstream" of the avidin-binding polypeptide (i.e. the identification protein)).

Thus, according to a first aspect of the present invention there is provided a recombinant hybrid polypeptide comprising a polypeptide of interest fused to an avidin-binding polypeptide containing a biotin attachment do-

main, characterised in that the polypeptide of interest is fused to the C terminus of the avidin-binding polypeptide.

Preferably, biotin is attached to the avidin-binding polypeptide.

5 Preferably, the polypeptide includes a cleavage site for cleaving the polypeptide of interest from the avidin-binding polypeptide. This cleavage site may be between the polypeptide of interest and the avidin-binding polypeptide or it may be integral with either or both of the polypeptide of interest and the avidin-binding polypeptide.

Preferably, the cleavage site is aspartic acid-proline, asparagine-glycine, methionine, cysteine, lysine-proline, arginine-proline, lysine-arginine or isoleucine-glutamic acid-glycine-arginine.

10 Preferably, the avidin binding polypeptide is, or is part of, a 1.3S polypeptide. Preferably, the 1.3S polypeptide is from *Propionibacterium*.

Preferably, the biotin attachment domain of the avidin-binding polypeptide comprises at least one of the sequence:

15 Pro Ala Pro Leu Ala Gly Thr Val Ser Lys Ile Leu Val Lys Glu
 Gly Asp Thr Val Lys Ala Gly Gln Thr Val Leu Val Leu Glu
 Ala Met Lys Met Glu Thr Glu Ile Asn Ala Pro Thr Asp Gly.

20 Preferably, the avidin-binding polypeptide comprises a plurality of non-contiguous and / or contiguous avidin-binding polypeptides, which may be the same or different.

Preferably, the polypeptide of interest comprises a plurality of non-contiguous and /or contiguous polypeptides of interest, which may be the same or different.

25 Preferably, the polypeptide of interest is an enzyme, is an antigen useful for vaccine production, or a diagnostic reagent.

Preferably, the polypeptide of interest has antitumor activity or has an amino acid sequence for recognition of antigens.

30 According to a second aspect of the present invention there is provided a nucleic acid sequence coding for a hybrid polypeptide comprising a polypeptide of interest and an avidin-binding polypeptide containing a biotin attachment domain, characterised in that the nucleic acid sequence coding for the polypeptide of interest is downstream of the nucleic acid sequence coding for the avidin-binding polypeptide.

Preferably, the nucleic acid sequence is a DNA sequence.

35 Preferably, the DNA sequence contains in a 5' to 3' direction on the coding strand a gene comprising a 5' promoter region, a DNA sequence coding for the avidin-binding polypeptide and the DNA sequence coding for a polypeptide of interest.

Preferably, the DNA sequence is, or is part of, an expression vector or a plasmid.

40 According to a third aspect of the present invention there is provided a process for the production of a hybrid polypeptide according to the first aspect of the present invention comprising constructing a plasmid containing a nucleic acid sequence according to the second aspect of the present invention, transforming the plasmid into a procaryotic or eucaryotic host cell expression system, expressing the system, contacting the hybrid polypeptide resulting from the expression system with avidin, and harvesting the resulting avidin-bound hybrid polypeptide.

Preferably, the expression system is either *E. Coli* or insect cells.

45 According to a fourth aspect of the present invention there is provided a process for the isolation of a hybrid polypeptide according to the first aspect of the present invention comprising contacting the hybrid polypeptide with avidin.

Preferably, the avidin is monomeric avidin, tetrameric avidin or streptavidin.

Preferably, the polypeptide of interest is cleaved from the isolated hybrid polypeptide.

50 Preferably, the hybrid polypeptide is isolated using avidin covalently bound to a chemically inert, solid, water and solvent insoluble substrate through a chemically stable non-hydrolyzable linking group.

Preferably the hybrid polypeptide is isolated using avidin monomer affinity chromatography. Preferably, the hybrid polypeptide is isolated using avidin monomer affinity chromatography disclosed in EP-A-0414785 (90310154.1).

55 EP-A-0414785 discloses a monomeric avidin polypeptide ligand and a novel and particularly efficacious process for isolating synthetic or natural molecules and / or biotinylated derivatives thereof, by adsorption of the molecules of interest onto a novel affinity media which contains avidin fixed to a solid inert support.

According to a fifth aspect of the present invention there is provided a first kit comprising a hybrid polypep-

tide according to the first aspect of the present invention and avidin.

According to a sixth aspect of the present invention there is provided a second kit comprising a nucleic acid sequence according to the second aspect of the present invention and avidin.

According to a seventh aspect of the present invention there is provided a third kit comprising a nucleic acid sequence which codes for an avidin-binding polypeptide containing a biotin attachment domain and which is fusable to a nucleic acid sequence coding for a polypeptide of interest in order to form a hybrid nucleic acid sequence according to the second aspect of the present invention and avidin.

Preferably, the third kit comprises means to fuse the nucleic acid sequence coding for the avidin-binding polypeptide to the nucleic acid sequence coding for the polypeptide of interest in order to form the nucleic acid sequence according to the second aspect of the present invention.

Preferably, any one of the kits comprises means to cleave the polypeptide of interest from the avidin-binding polypeptide.

In accordance with the present invention, therefore, a hybrid polypeptide comprising one or more polypeptides of interest and at least one polypeptide for attachment is produced by recombinant DNA methods.

The present invention also provides a process for producing this hybrid polypeptide in a procaryotic or a eucaryotic protein expression system.

One of the advantages of the present invention is that it provides a means for isolating a hybrid polypeptide in high purity and yield.

Another advantage is that the hybrid polypeptide according to the present invention can be recovered in high purity and high yield in a single chromatographic step, such as the avidin monomer affinity chromatography technique disclosed in EP-A-0414785.

Another advantage of the present invention resides in the fusion of the polypeptide of interest to the C-terminus, and not to the N-terminus, of the polypeptide containing the biotin domain to avidin (e.g. a 1.3S polypeptide).

In this regard, the protein expression level in a host cell is determined by a number of factors, including promoter strength and optimal initiation of protein translation (see commentary above). Promoter strength contributes to the efficiency of transcription of messenger RNA. Optimization of the processes involved in the initiation of translation is important to achieving high levels of protein expression in the host cell. When polypeptides of interest are introduced at the 3' terminus of the gene coding for the polypeptide containing the biotin domain to avidin (e.g. a 1.3S gene), no change is made to the optimal placement of the 5' terminus of the 1.3S gene directly adjacent to the promoter and 5' regulatory sequences. Thus, maximal expression levels in host cells can be achieved.

In contrast to this, and as found with the prior art methods, if the polypeptide of interest is inserted between the promoter and 5' terminus of, for example, the 1.3S gene additional experimentation and tailoring is required to achieve maximal expression levels in host cells.

The fusing the polypeptide of interest to the C terminus of the polypeptide containing the biotin domain to avidin (e.g. a 1.3S polypeptide or a fragment of the 1.3S polypeptide), so that the correct conformation of the biotin attachment region may be preserved, is thus in direct contrast to the prior art methods (such as those disclosed in Murtif and Samols (ibid), PCT WO 90/14431, Cronan (ibid), US-A-4782137 and US-A-4839293, which in fact teach away from the present invention).

Moreover, it was surprising to find that if one went against the teachings of the prior art methods (such as those disclosed in Murtif and Samols (ibid), PCT WO 90/14431, Cronan (ibid), US-A-4782137 and US-A-4839293) and fused a polypeptide of interest to the C-terminus of the polypeptide containing the biotin domain to avidin, such as a 1.3S polypeptide, the appropriate lysine residue of the 1.3S polypeptide within the hybrid was indeed biotinylated. This is surprising since, if one were to follow the earlier teachings of Murtif and Samols, the addition of a polypeptide substantially longer than two amino acid residues would be expected to alter the conformation of the C terminus of the 1.3S polypeptide and thus preclude biotinylation.

It is also surprising to find that, contrary to the teachings of Cronan (ibid) and Murtif and Samols (ibid), when a polypeptide of interest is fused to the polypeptide containing the biotin domain to avidin, such as a 1.3S polypeptide, at the C-terminus of the 1.3S polypeptide, biotin is attached to the biotin-attachment domain of 1.3S polypeptide within this hybrid recombinant polypeptide. In this regard, Cronan (ibid) and Murtif and Samols (ibid) teach that fusion at the C-terminus of the 1.3S may disrupt the native hydrophobicity and thus the native conformation of the 1.3S polypeptide, thus inhibiting biotinylation, and consequently inhibiting the binding of hybrid polypeptides to avidin.

It is further surprising to find that the biotin group attached to the polypeptide containing the biotin domain to avidin, such as a 1.3S peptide, fused at its C-terminus to another polypeptide is positioned so as to make the biotin molecule available for binding the hybrid polypeptide to the avidin monomer affinity resin of EP-A-0414785.

Furthermore, the hybrid polypeptide is selectively retained by the avidin resin and can be recovered in high yield and high purity.

In the present invention, a binding domain, or recognition sequence, directs the attachment of biotin to the hybrid polypeptide. The biotinylated hybrid polypeptide can then be specifically selected by affinity ligand compositions such as the avidin monomer resin of EP-A-0414785.

The present invention therefore yields a single purification step that can separate the protein of interest from complex mixtures, such as crude bacterial lysates, with high levels of recovery in a single chromatographic step, thus alleviating the recovery problems inherent to multistep purification processes. Such a combination of hybrid polypeptide and avidin monomer affinity resin would clearly confer significant advantages to the purification of commercially useful polypeptides over existing processes.

An avidin-binding polypeptide for attachment is generally a polypeptide that enables the attachment of a hybrid fusion polypeptide to avidin. Among such polypeptides for attachment that may be used are those containing a recognition sequence for attachment of the prosthetic group biotin, such as the 1.3S polypeptide subunit of transcarboxylase from *Propionibacterium*.

In addition to using the entire sequence of the 1.3S polypeptide as the polypeptide for attachment, other smaller portions of the 1.3S polypeptide may be used which direct the attachment of biotin, particularly portions comprising all or part of amino acid residues 58 through 100 (SEQ ID NO: 2).

One or more deletions, substitutions, insertions or mutations may be made by methods well known in the art which result in a biotinylated 1.3S polypeptide or biotinylated fragment. The nucleotide sequence coding for the 1.3S polypeptide or fragments may be synthesized using a commercially available DNA synthesizer in a manner well known in the art.

Additionally, other polypeptides or portions thereof that are enzymatically biotinylated may also be employed.

Unless indicated otherwise, the term "avidin" includes streptavidin.

The polypeptide of interest may include two or more polypeptides of interest.

The two or more polypeptides of interest may be fused sequentially.

Optionally, contiguous polypeptides of interest may be fused sequentially.

Additionally, more than one polypeptide of interest may be present in a noncontiguous arrangement, for example, one polypeptide of interest may be fused to the N-terminus of the polypeptide for attachment and one polypeptide fused at the C-terminus of the same polypeptide for attachment.

Two polypeptides of interest may be fused to the C-termini of two polypeptides for attachment arranged as: polypeptide for attachment 1 - polypeptide of interest 1 - polypeptide for attachment 2 - polypeptide of interest 2.

In any of the arrangements disclosed here, the polypeptides of interest may be the same, or different. The polypeptides for attachment may be the same, or different.

An advantage of fusing a plurality of polypeptides of interest to at least one polypeptide for attachment is the ability to increase the yield of a single polypeptide of interest present by including two copies of that polypeptide within a single hybrid polypeptide, and/or to increase the number of polypeptide species that can be purified simultaneously by a single avidin affinity chromatography step, if the polypeptides of interest are different.

A cleavage amino acid or sequence of amino acids may be present between the polypeptide containing the biotin domain to avidin and the polypeptide of interest.

Likewise, a cleavage amino acid or sequence of amino acids may be present between the polypeptides of interest if two or more of such polypeptides are present. Such linking, or cleavage, amino acid(s) permits the separation of polypeptides at a specific site or sites on the hybrid polypeptide when it is treated with the appropriate chemical reagent or enzyme. If desired, the cleavage site is positioned adjacent the polypeptide of interest so that the polypeptide of interest may be cleaved from the polypeptide for attachment.

The hybrid polypeptide itself may contain a linking amino acid or amino acids for cleaving the polypeptide or polypeptides of interest from the polypeptide or polypeptides for attachment. The linking amino acid or amino acids are incorporated between the polypeptide or polypeptides for attachment and the polypeptide or polypeptides of interest in such a way that one or more cleavage reactions separate each polypeptide species to the degree necessary for intended applications. It may not in every instance be necessary to cleave all, some, or any of the species within a particular hybrid polypeptide.

Amino acids that may be used to link the polypeptide of interest to the polypeptide for attachment include aspartic acid-proline, asparagine-glycine, methionine, cysteine, lysine-proline, arginine-proline, isoleucine-glutamic acid-glycine-arginine, and the like.

The at least one polypeptide for attachment may be cleaved from the at least one polypeptide of interest by exposure to the appropriate chemical reagent or cleaving enzyme.

It should be recognized that cleavage of the polypeptide or polypeptides of interest from the polypeptide or polypeptides for attachment may not be necessary for every hybrid fusion polypeptide that is constructed, in which case a cleavage site could be incorporated, or absent.

5 The polypeptide of interest can comprise substantially any procaryotic or eucaryotic polypeptide that can be expressed by a vector in a host cell. Among the polypeptides of interest which may be produced by such means are enzymes, such as proteases, oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases.

10 The present invention also contemplates the production of storage polypeptides, such as ferritin or ovalbumin or transport polypeptides, such as hemoglobin, serum albumin, eruloplasmin, or the like. Also included are the types of polypeptides that function in contractile and motile systems, for example actin and myosin or the like.

The present invention also contemplates the production of polypeptides that serve a protective or defense function, such as the blood polypeptides thrombin and fibrinogen.

15 Other protective polypeptides include the binding polypeptides, such as antibodies or immunoglobulins that bind to and thus neutralize antigens. Additionally this invention contemplates Protein A, or the like.

The polypeptide produced by the present invention also may encompass various hormones such as endorphins, human growth hormone, somatostatin, prolactin, estrogen, progesterone, thyrotropin, calcitonin, gonadotropin, insulin or the like.

20 Other such hormones include those that have been identified as being involved in the immune system, such as interleukin 1, interleukin 2, colony stimulating factor, macrophage-activating factor, interferon, or the like.

The present invention may be used to produce toxic polypeptides, such as ricin from castor bean or gossypin from cotton seed, and the like.

Polypeptides that serve as structural elements may be produced by the present invention, such polypeptides include the fibrous polypeptides collagen, elastin and alpha-keratin.

25 Other structural polypeptides include glycoproteins, virus-proteins, muco-proteins and the like.

Polypeptides that may be utilized as diagnostic agents, for example as markers for the presence of certain diseases, are also contemplated by this invention.

30 Additional polypeptides of interest that may be produced as hybrid polypeptides are polypeptides that may be used for therapeutic purposes, for example polypeptides with anti-tumor activity, polypeptides useful in vaccine production, polypeptides having amino acid sequences for recognition of antigens, or polypeptides which can function as diagnostic reagents, and the like.

In addition to the above-noted naturally occurring polypeptides, the present invention may be used to produce synthetic polypeptides, defined generally as any sequence of amino acids not occurring in nature.

35 Preferably, the hybrid polypeptide is produced in procaryotic or eucaryotic cells transformed by a cloning vector comprising a nucleic acid sequence according to the present invention. The hybrid polypeptide is then purified away from the complex cell extract mixture by avidin affinity chromatography. A particularly preferred form of avidin is avidin monomer.

40 In a preferred embodiment, an extract of transformed cells is made from cell culture or fermentation broth, the hybrid polypeptide is then rendered to a soluble state, and the extract is then applied to the avidin monomer column. The column is then washed with adequate amounts of a wash buffer to clear the column of unbound materials. The hybrid polypeptide is then eluted from the column.

45 After the hybrid polypeptide is eluted from the column, the polypeptide for attachment may optionally be cleaved from the polypeptide of interest with the appropriate cleavage reagent or enzyme. Passage of the cleaved mixture over the avidin monomer column yields a highly purified preparation of the polypeptide of interest, and the polypeptide for attachment is then retained by the column.

The avidin used to bind the biotin attached to the hybrid polypeptide may be monomeric or tetrameric avidin, or streptavidin. Avidin monomer is the preferred form of avidin affinity medium.

50 Advantages of using avidin monomer to separate the polypeptide of interest from crude cell mixtures include reversible binding of the polypeptide for attachment to avidin, high yield, and high purity of the desired polypeptide of interest following affinity chromatography.

The genes coding for hybrid polypeptides may be produced by recombinant DNA methods by combining within a DNA expression vector a chimeric gene comprising a 5' promoter region, DNA sequences coding for at least one polypeptide for attachment of a prosthetic group for binding to avidin, and at least one DNA sequence coding for a polypeptide of interest.

55 Optionally, the chimeric gene may contain at least one DNA sequence coding for a linking amino acid or amino acids, that is, one or more amino acids for cleaving a polypeptide of interest from a polypeptide for attachment.

Genes coding for the various types of polypeptides of interest, for example those identified above, may be

obtained from a variety of procaryotic or eucaryotic sources, such as plant or animal cells or bacterial cells. Genes can be isolated from chromosomal material of eucaryotic or procaryotic cells, or from plasmids or viruses of procaryotic or eucaryotic cells by employing standard, well-known techniques.

5 Additionally, automated DNA synthesis may be used to obtain DNA coding for naturally-occurring or synthetic polypeptides. To enable chimeric gene expression in host cells, a variety of naturally-occurring and synthesized DNA expression vectors having genes coding for many different polypeptide molecules are now commercially available from a variety of sources.

The desired DNA can also be produced from mRNA by using the enzyme reverse transcriptase. This enzyme permits the synthesis of DNA from an RNA template.

10 In accordance with the present invention, once genes coding for one or more desired polypeptides of interest are isolated, synthesized or otherwise obtained, said gene or genes are joined to at least one gene coding for a polypeptide containing a recognition sequence for attachment of the prosthetic group biotin, thus enabling the attachment of the hybrid polypeptide to avidin.

A gene directing the synthesis of a polypeptide for attachment is generally one coding for a polypeptide that enables the binding of a hybrid fusion polypeptide to avidin. Among such genes coding for polypeptides for attachment that may be used are those coding for amino acid sequences that direct the attachment of the prosthetic group biotin, for example the gene for the 1.3S polypeptide subunit of transcarboxylase from *Propionibacterium*.

20 A gene coding for a polypeptide for attachment may be one that directs the attachment of the prosthetic group biotin. A particularly biotinylated preferred biotinylated polypeptide for attachment is the 1.3S subunit of transcarboxylase from *Propionibacterium shermanii* (SEQ ID NO:1). Although the gene coding for the entire 1.3S polypeptide from *Propionibacterium shermanii* is preferred (SEQ ID NO: 4), optionally any gene or gene fragment coding for a polypeptide that directs the attachment of biotin may be suitable. The preparation of gene fragments is well known to those skilled in the art.

25 The gene or genes coding for the at least one polypeptide of interest and the gene or genes coding for the polypeptide or polypeptides for attachment are preferably treated with the appropriate restriction enzymes, or otherwise treated to have cohesive termini to facilitate ligation with other elements of the chimeric gene or the DNA expression vector.

30 The resulting DNA expression vector carrying the chimeric hybrid polypeptide gene is used to transform the appropriate procaryotic or eucaryotic host cell. The selection of a DNA expression vector appropriate for the desired host cell is well known to those skilled in the art.

35 Following the transformation procedure, the transformed host cells are isolated and analyzed for expression of the hybrid polypeptide. Those transformants identified as containing the hybrid polypeptide are further analyzed by restriction enzyme digestion, DNA sequencing and other methods for confirming the correctness of the desired gene, by methods well known to those skilled in the art.

The transformants identified as host cells carrying the gene for the desired hybrid polypeptide are then multiplied in culture to cause replication of the vector and high-level expression of the hybrid polypeptide that contains the polypeptide of interest.

40 The cloning vector may be used to transform additionally other strains of compatible hosts for large-scale production of the hybrid polypeptide.

45 Various methods used for obtaining genes or gene fragments, preparing DNA expression vectors, transforming host cells, expressing hybrid polypeptides in host cells, and identifying those polypeptides are set forth by J. Sambrook, E.F. Fritsch, and T. Maniatis, *Molecular Cloning*, 2nd Edition. Cold Spring Harbor Press, 1989, and also by F.M. Ausubel, R. Brent, R.E. Kingston, D.M. Moore, J.G. Seidman, J.A. Smith, K. Struhl., Eds. *Current Protocols in Molecular Biology*, Volume 1. John Wiley and Sons, New York 1989.

50 To prepare DNA expression vectors, various cloning vectors may be used. A plasmid is preferred. However, a cosmid or bacteriophage may be used. If insect, plant, or mammalian cells are used as host cells, viruses may also be used as vectors. DNA expression vectors may be obtained from natural sources or may comprise synthetic DNA. The plasmid chosen for a particular expression system should be compatible with that host, to ensure vector replication and polypeptide expression. The plasmid chosen for incorporation of the genes coding for the hybrid polypeptide should possess an origin of replication recognized by the host cell.

55 The DNA expression vector should contain DNA sequences recognized by restriction endonuclease enzymes to cleave the vector for subsequent ligation with the gene for the hybrid polypeptide without inactivating the origin of replication or functions necessary for plasmid selection following transformation, for example within an antibiotic resistance gene. The vector should contain restriction enzyme cleavage sites that provide suitable termini for joining and ligation of foreign genes to be inserted.

Preferably, the DNA vector contains a single site or two unique sites for incorporation of the hybrid polypeptide gene, neither of which occurs within that gene.

To accommodate more than one different foreign gene possibly terminating in different cohesive or blunt termini, it would be useful for the vector to possess a large number of unique restriction enzyme cleavage sites.

Preferably, the DNA expression vector will cause a phenotype to be expressed that will enable transformed cells to be readily identified and separated from cells which do not undergo transformation. Such phenotypic selection genes can include genes providing resistance to a particular antibiotic, which inhibits growth of untransformed but not transformed cells. Such genes are widely available now and confer resistance to antibiotics such as ampicillin, tetracycline, streptomycin, kanamycin, and the like.

Plasmids which contain an inserted gene that disrupts the B-galactosidase gene, such as the hybrid polypeptide gene, can be identified following transformation by the inability of the host cell to reduce reagent 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-gal) in the media and cause the bacterial colony to develop a blue coloration. Such plasmids, reagents, and media are known to those skilled in the art.

Preferably, *E. coli* is employed as the host cell, and a plasmid is preferred for cloning and transformation of the *E. coli* host. The preferred plasmid is pKK223-2 (Pharmacia, Uppsala, Sweden). This plasmid carries genes for an origin of replication in *E. coli*, and a gene for resistance to the antibiotic ampicillin. This plasmid also has a synthetic linker region consisting of unique restriction endonuclease cleavage sites to facilitate cloning. This plasmid contains the strong *tac* promoter, which directs high levels of transcription in *E. coli*.

If insect cell culture is to be used for production of hybrid polypeptide, the preferred plasmid is pVL1392 (obtained from M. Summers, University of Texas, available commercially from Invitrogen, Inc. San Diego CA).

An advantage of insect cell culture is that polypeptides requiring glycosylation or other types of post translational modifications including folding with appropriate disulfide bond formation may be so modified in an insect cell expression system, whereas this manner of post-translational modification is not performed by procaryotic hosts.

To prepare the chosen plasmid for insertion of the chimeric gene comprising the hybrid polypeptide gene, the plasmid is digested with restriction endonucleases, for example BamHI or EcoRI, or any restriction enzyme or enzyme combination that cleaves the plasmid at a unique site and produces cohesive 3' and 5' termini complementary to termini of the chimeric gene to be ligated. If desired, the plasmid may be treated with two different enzymes to produce two different cohesive termini to facilitate ligation of the chimeric hybrid polypeptide gene or genes in the correct orientation within the plasmid. Certain enzymes which produce blunt ends may also be used, or linker molecules may be added to vector or foreign genes to prepare the desired cohesive termini. Such strategies and methods are well known to those skilled in the art.

When the plasmid is digested, two or more DNA fragments may be generated. The desired plasmid fragment carrying the origin of replication and other genes essential to replication and identification of the plasmid may be identified and recovered by gel electrophoresis and other techniques well known in the art.

A particularly preferred arrangement for the members of the hybrid polypeptide is the location of the polypeptide for attachment, most preferably the 1.3S polypeptide directly 3' to the promoter at the PstI site within the synthetic linker of the plasmid pkk223-3. The PstI site is 3' to the promoter and to a ribosome binding site.

It should be understood that any deletions, insertions, substitutions, or mutations which may be performed on the 1.3S gene which still direct the attachment of biotin are contemplated within the spirit and scope of this invention. Additionally, other genes or gene fragments, natural or synthetic, whose resulting polypeptides direct the attachment of biotin or lipoic acid fall within the scope and spirit of this invention.

The 1.3S gene is preferably constructed with a cleavage site for PstI at its 5' terminus and a cleavage site for BamHI at its 3' terminus, such that upon ligation, the 1.3S gene is connected in the proper reading frame with the *tac* promoter, a ribosome binding site is intact, and the 1.3S gene or fragment preferably terminates in the nucleotide sequence GAT CCA TAA CGC CTA AGC TT (SEQ ID NO: 3), or any such sequence which simultaneously provides a BamHI restriction endonuclease cleavage site and codes for the amino acids asp-pro. Asp pro is the preferred sequence used as linking amino acids the cleavage of the 1.3S polypeptide for attachment from appropriate polypeptides of interest.

However, if a polypeptide of interest contains within its sequence one or more asp-pro sequences, then optionally any other linking amino acid or amino acids not present in the polypeptide of interest may be substituted for asp-pro. It will be necessary to structure the gene in such instances that appropriate cohesive termini are created that permit ligation of the 1.3S gene to the gene for the polypeptide of interest.

The gene or genes for the at least one polypeptide of interest may be isolated, synthesized, or otherwise obtained and modified at the 5' terminus so that ligation to the appropriate terminus of the gene for the polypeptide for attachment is facilitated. The 3' terminus of the 1.3S polypeptide for attachment is the preferred terminus for ligation of the gene for the polypeptide of interest, in the proper reading frame.

Furthermore, the 3' terminus of the last polypeptide of interest in sequence in the chimeric gene should preferably be prepared so that this terminus is complementary to the 5' terminus of the plasmid vector, to facilitate ligation to the expression vector.

It is to be understood that for different chimeric genes, obtaining the correct orientation of the polypeptide or polypeptides of interest relative to each other and to the polypeptide or polypeptides for attachment within the expression vector employ the basic steps as outlined above.

5 Preferably, the gene for each polypeptide of interest is attached in the proper reading frame to the adjacent gene, and all adjacent termini are prepared in a manner so as to make them complementary to facilitate ligation.

Furthermore, genes for any polypeptides of interest requiring cleavage from another polypeptide of interest or from one or more polypeptides for attachment must be so constructed as to allow for the proper positioning of all cleavage sites so that their insertion does not result in any genes for any of the polypeptides being in an improper reading frame.

10 The ligation reaction, which covalently joins fragments of DNA, is described in Sambrook, Fritsch, and Maniatis (ibid), and Ausubel et al. (ibid) and is well known to those skilled in the art.

The ligated plasmid is ready for transformation of host cells. The preferred host is *E. coli*, however, other bacteria, insect cells, yeast, or mammalian or plant cells may be used with a DNA expression vector appropriate to that particular host cell.

15 Transformation of *E. coli* is a standard procedure well known to those skilled in the art, wherein a suitable host strain, such as *E. coli* HB101 accepts, harbors, replicates and expresses the plasmid carrying the gene for the hybrid polypeptide. Transformation of *E. coli* is described by Sambrook, et al. If the host is an insect cell, transfection may be accomplished by a procedure such as that described by M. Summers and G. Smith, A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures. Texas Agricultural Experiment Station Bulletin No. 155, 1988.

In order to identify the host cells which are transformed, the culture is placed in selective media containing an appropriate antibiotic. Only those cells with plasmid-borne resistance will survive.

25 Plasmid can be recovered after lysis of surviving cell colonies, and characterized by restriction enzyme digestion and mapping, DNA sequencing, or other methods known in the art. Additionally, those colonies which express hybrid polypeptide can be identified by immunological assay, such as ELISA or Western blotting. In some embodiments it may be possible to assay directly for biological activity of the polypeptide or polypeptides of interest.

30 Once transformed cells carrying the hybrid polypeptide are identified, they may be multiplied by established techniques, such as fermentation. In addition, the recovered plasmids can be used to transform other strains of bacteria, or appropriate hosts cells for large-scale production and expression of the hybrid polypeptide.

The hybrid polypeptide which contains the polypeptide for attachment, the biotin group for binding to avidin, the polypeptide of interest, and the optional cleavage site, expressed by the transformed host cells may be separated from the medium and other debris by affinity chromatography.

35 The preferred affinity medium is the avidin monomer resin described in EP-A-0414785. To this end, host cells are separated from the medium and broken open for example, by sonication.

Optionally, hybrid polypeptides can be excreted into the culture media if a signal peptide for extracellular secretion is included at the appropriate terminus of the hybrid polypeptide.

Should such a secreted polypeptide be desired, it may be necessary to include a DNA sequence coding for a polypeptide directing extracellular secretion within the chimeric gene coding for the hybrid polypeptide.

40 The hybrid polypeptide once released is maintained in an appropriate buffer, preferably one in which it is soluble. The buffer solution should be formulated to maximize hybrid polypeptide recovery from host cells. Buffer properties which may be optimized to favor recovery include but are not limited to, pH, ionic composition, ionic strength, or presence or absence of various detergent compositions.

45 Optionally some fractionation of the host cell extract may be performed in order to concentrate or partially purify the hybrid polypeptide prior to affinity chromatography.

One preferred method is ammonium sulfate fractionation. It is to be understood that other methods commonly employed in protein purification may also be used prior to affinity chromatography. The cell extract is passed over the preferred column for affinity chromatography, the avidin monomer column, which is then washed extensively with buffer to remove all unbound materials.

50 The hybrid polypeptide is specifically eluted from the column, preferably with acetic acid or biotin. As a result, a high yield of highly purified hybrid polypeptide containing the polypeptide of interest is obtained.

It may be desirable or necessary to cleave the one or more polypeptide of interest from the one or more polypeptide for attachment to restore biological activity to the polypeptide of interest.

55 Separation from the polypeptide for attachment may be accomplished by first suspending the hybrid polypeptide in buffer. Thereafter the chemical or proteolytic cleavage agent specific to the linking amino acid or amino acids is added to the suspension and the polypeptide of interest is cleaved.

For example, if the polypeptide of interest is linked to the polypeptide for attachment by an asp-pro linkage, a volatile acid such as formic acid may be added to the suspension to effect cleavage.

If methionine is the linking amino acid, the reagent cyanogen bromide may be used to cleave between methionine and the first amino acid of the polypeptide of interest.

A volatile cleavage reagent, such as formic acid or cyanogen bromide, may be evaporated away from the polypeptide mixture. If cleavage is accomplished by an enzyme, the enzyme may be removed from the mixture by passing the mixture through an enzyme substrate column.

If it is necessary to obtain the polypeptide of interest pure from the polypeptide for attachment, this removal may be accomplished by passing the mixture through an avidin affinity column. In this way, the polypeptide for attachment, by binding to the avidin, will be retained and therefore separated from the highly purified solution of the polypeptide of interest, which will not bind to avidin.

It should be noted that some polypeptides of interest will assume their desired biological activity with the polypeptide for attachment still attached. As a consequence, the polypeptide for attachment will not need to be cleaved from the polypeptide of interest and the steps described to separate the polypeptide of interest and the polypeptide for attachment need not be performed. Moreover, in circumstances where the polypeptide for attachment remains attached to the polypeptide of interest, linking amino acid or amino acids may be present or omitted. In this situation, the construction and method of preparing the DNA expression vectors, detailed above, can be appropriately modified.

The present invention will now be described by way of examples only.

Reference shall be made to the following figures, in which:

Figure 1 shows a partial restriction map of plasmid ptac1.3dp;

Figure 2 shows chimeric gene constructs for hybrid polypeptides constructed so that the polypeptide of interest is fused at the C-terminus of the polypeptide for attachment;

Figure 3 shows chimeric genes for hybrid polypeptides containing more than one polypeptide of interest fused to a single polypeptide for attachment; and

Figure 4 shows chimeric gene constructs for a hybrid polypeptide containing two noncontiguous polypeptides of interest, each fused to the C-terminus of noncontiguous polypeptides for attachment.

Referring to Figure 1, a partial restriction map of plasmid ptac1.3dp is shown. ptac1.3dp was created by modification of plasmids ptac1.3t and ptac1.3(1-125) obtained from D. Samols, Case Western Reserve University. An E. Coli strain CSH26 containing the plasmid ptac1.3dp has been deposited in the American Type Culture Collection, Rockville, Md, USA as ATCC No. 68937. This deposit was made pursuant to the Budapest Treaty On The International Recognition Of The Deposit Of Microorganisms For The Purposes Of Patent Procedure.

Referring to Figure 2, chimeric gene constructs for hybrid polypeptides are shown. The chimeric genes are constructed so that the polypeptide of interest is fused at the C-terminus of the polypeptide for attachment. In these examples, the polypeptide of interest is a synthetic b-endorphin, and the polypeptide for attachment is the 1.3S polypeptide from transcarboxylase of *Propionibacterium shermanii*.

In Figure 2A is shown an asp-pro cleavage site located between the C-terminus of the 1.3S polypeptide and the N-terminus of the b-endorphin polypeptide.

In Figure 2B is shown an asp-pro cleavage site located between the C-terminus of the 1.3S polypeptide and the N-terminus of a novel reverse-endorphin polypeptide.

In Figure 2C is shown a methionine cleavage site located between the C-terminus of the 1.3S polypeptide and the N-terminus of the b-endorphin polypeptide.

In Figure 2D, no cleavage site is located between the C-terminus of the 1.3S polypeptide and the N-terminus of the b-endorphin polypeptide.

Referring to Figure 3, chimeric genes for hybrid polypeptides are illustrated which contain more than one polypeptide of interest fused to a single polypeptide for attachment.

In Figure 3A is shown the fusion of two contiguous b-endorphin polypeptides to the C-terminus of the 1.3S polypeptide from transcarboxylase of *Propionibacterium shermanii*.

In Figure 3B is illustrated the fusion of two different noncontiguous polypeptides of interest to the 1.3S polypeptide. A maltose binding protein is fused to the N-terminus of the 1.3S polypeptide, and a synthetic b-endorphin is fused to the C-terminus of the same 1.3S polypeptide.

Figure 3C shows the fusion of two different contiguous polypeptides of interest to the N-terminus of the 1.3S polypeptide.

The maltose binding protein and a synthetic b-endorphin polypeptide are fused in tandem to the N-terminus of the 1.3S polypeptide.

Referring to Figure 4, chimeric gene constructs for a hybrid polypeptide containing two noncontiguous polypeptides of interest are shown. Each coding sequence is fused to the C-terminus of a noncontiguous polypeptides for attachment. In the specific example, two synthetic b-endorphin polypeptides are each fused to the C-terminus of a different 1.3S polypeptide.

The following examples are carried out using one or more of the general procedures set forth below.

In all of the following examples, restriction endonucleases, ligases, polymerases, and other DNA modifying enzymes described in specific experimental steps are used according to the recommendations of the manufacturer of the particular enzyme or reagent used.

- 5 Two laboratory manuals, Current Protocols in Molecular Biology (Ausubel et al 1989)), and Molecular Cloning A Laboratory Manual 2nd Edition (Sambrook et al, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)) referenced below, contain supplemental information that may also be useful to one skilled in the art in conducting the examples described below.

10 **Procedure I. Restriction Endonuclease Digestion of DNA**

Restriction endonuclease digestions using one or more restriction enzymes to digest DNA are generally carried out using the protocols set forth in Ausubel et al. Current Protocols in Molecular Biology Volume 1, Chapter 3, Unit 3.1.

- 15 Restriction mapping of plasmids is generally carried out using the protocols set forth in Ausubel et al (ibid), Unit 3.2. Restriction enzymes are obtained from Promega (Madison WI) or New England Biolabs (Beverly MA) and complete or partial digestion of DNA with specific enzymes are performed generally according to the manufacturer's recommendations .

20 **Procedure II. Purification of DNA Fragments Using Agarose Gel Electrophoresis.**

- Agarose gel electrophoresis is generally carried out using the protocols set forth in Ausubel et al (ibid), Volume I, Chapter 2, Unit 2.5A. Separation and isolation of larger (>1 kb) DNA fragments from excised gel fragments is generally carried out as set forth by Ausubel et al (ibid), Chapter 2, Unit 2.6, and separation and isolation of smaller (<1 kb) DNA fragments is carried out as set forth in Ausubel et al (ibid) Chapter 2, Unit 2.7. Removal of salts and gel fragments from larger DNA fragments is accomplished using the GeneClean kit (Bio101, Inc. San Diego, CA) using procedures supplied by the manufacturer, and removal of salts and gel fragments from smaller DNA fragments is accomplished using the MerMaid kit (Bio101, Inc. San Diego, CA) also using the procedures recommended by the manufacturer.

30 **Procedure III. Ligation of DNA Fragments.**

Ligation of DNA fragments using T4 DNA Ligase (New England Biolabs Beverly MA) is generally carried out as in Ausubel et al (ibid), Unit 3.14, using conditions recommended by the manufacturer.

35 **Procedure IV. Preparation of Competent E. coli Cells and Transformation of E. coli.**

- Preparation of competent E. coli CSH26 cells using calcium chloride and transformation with DNA expression vectors are carried out using the protocols described by Sambrook et al Molecular Cloning A Laboratory Manual 2nd Edition 1989, Chapter 1.

Clones carrying plasmids containing DNA insertions are identified by growing cells on L-agar supplemented with 100 mg/L ampicillin.

45 **Procedure V. Isolation of Plasmid DNA.**

Plasmid DNA is generally usubel et al (ibid) Chapter 1, Unit 1.7.

Procedure VI. Preparation of Double-stranded DNA from Synthetic Single-stranded Oligonucleotides.

- 50 Oligonucleotides are synthesized using standard phosphoramidite chemistry (0.2 micromole synthesis). Fragments are separated on a 20% polyacrylamide gel under denaturing conditions as described by Ausubel et al (ibid) Volume 1, Unit 2. 12 and eluted and desalted as described therein.

- Double-stranded DNA is assembled from upper strand and lower strand pairs of synthetic oligonucleotides with overlapping regions of perfect complimentarity of 15 nucleotides at their 3' ends by heating a mixture of 1 ug of each of the strands at 90°C for 5 minutes, followed by slow cooling to room temperature over a period of one to two hours. This short duplex region serves as template and primer for mutually primed synthesis of a complete double DNA strand with Sequenase, a T7 DNA polymerase obtained from US Biochemical, using protocols supplied by the manufacturer.

Following duplex extension, the DNA double strand is purified by agarose gel electrophoresis as described in Procedure II.

Procedure VII. Preparation of Crude E. coli Cell Extract.

E. coli host cells harboring the plasmid carrying the gene for the hybrid polypeptide are grown to stationary phase by overnight incubation in L-broth containing 100 mg/L ampicillin at 42°C, 250 RPM in a New Brunswick incubator-shaker.

Cells are collected by centrifugation in a GSA rotor at 5,000 X G for 30 minutes at 4°C in a Sorvall RC-5B centrifuge to pellet the cells. The supernatant is poured off, and the cell pellet is weighed. The cells are resuspended in 1:2 (cell wet weight: buffer volume) 100 mM potassium phosphate buffer, pH 6.8-7.2 (or appropriate buffer at pH 4.0 to 11.0), at 4°C.

The cells are then lysed by sonication using a large probe on Fisher sonic dismembrator Model 300 for 3 one minute cycles at 95% relative output. The resulting lysate is centrifuged at 17,500 RPM for 30 minutes at 4°C. To the resulting supernatant is added 2% (w:v) streptomycin sulfate. After incubation for 15 to 30 minutes at 4°C, the lysate is centrifuged at 17,500 RPM as above. The resulting supernatant is adjusted to 30% saturation with ammonium sulfate, incubated for 30 minutes at 4°C, and centrifuged at 17,500 RPM as above. The supernatant is adjusted to 60% saturation with ammonium sulfate, incubated for 30 minutes at 4°C, then centrifuged at 17,500 RPM as above.

The pellet formed by the addition of 60% ammonium sulfate is resuspended in 100 mM potassium phosphate buffer, pH 7.0 to yield approximately 80 to 100 mg/ml total protein, and is centrifuged at 17,500 RPM as above. The supernatant is termed the crude extract.

Procedure VIII. Avidin Monomer Affinity Chromatography.

The crude extract is applied to a 4mm X 5 cm column packed with avidin monomer affinity resin (US serial no. 414,785) on a LKB HPLC system equipped with two Model 2150 pumps, Model 2152 controller, and a Model 2140 spectral detector. Sample absorbance is monitored at 280 nm. The crude extract is applied in 100 mM potassium phosphate buffer, pH 6.8 to 7.2 (or other appropriate buffer at a pH of 4.0 to 11.0) at a flow rate of 0.1 ml /min to a column equilibrated with the same buffer.

After the sample is loaded, the column is washed with phosphate buffer at a flow rate of 1 ml /min until absorbance returned to the baseline absorbance, and all nonbound material is washed from the resin. The column is then equilibrated with water, followed by application of 5 ml of 2M NaCl. The column is reequilibrated with water. This same NaCl-water wash procedure is repeated four to five times. The sample is eluted using acetic acid or biotin, as detailed in Procedure IX.

Procedure IX. Elution of the Hybrid Polypeptide from the Avidin Monomer Affinity Resin

A. Elution using acetic acid

Five ml of 10% glacial acetic acid are applied to the column. Eluted hybrid polypeptide is collected until the absorbance at 280 nm returns to the baseline absorbance.

B. Elution using biotin

Five ml of 10 mM biotin in 100 mM potassium phosphate buffer, pH 6.5 is applied to the column. Eluted hybrid polypeptide is collected until the absorbance at 280 nm returns to the baseline absorbance.

Procedure X. Cleavage of the Polypeptide of Interest from the Polypeptide for Attachment

A. Acid cleavage

[Reference: London, M. (1977) Methods in Enzymology 47:145-149.]

The hybrid polypeptide suspension is adjusted to 70% formic acid (v/v) and incubated at 40°C for 24 to 48 hours. The mixture is then freeze-dried. Highly pure polypeptide of interest is obtained by Procedure VIII.

B. Cyanogen bromide cleavage of methionine residues [Reference: Gross, E. and B. Witkop (1961) journal of American Chemical Society 83:1510-1511.]

The hybrid polypeptide is dissolved in 70% (v/v) aqueous formic acid at 23°C. A 50 molar excess of cyanogen bromide is added in a small volume of 70% formic acid, with stirring.

The mixture is incubated in the dark under nitrogen at 20-25°C for 16 to 24 hours. The mixture is then diluted with 10 volumes of water and freeze-dried.

Highly pure polypeptide of interest is obtained by Procedure VIII.

Procedure XI. Separation of the Polypeptide of Interest from the Polypeptide for Attachment

The dried polypeptide mixture is resuspended in avidin monomer column loading buffer, 100 mM phosphate buffer pH 6.8-7.2, or other appropriate buffer at a pH of between 4.0 and 11.0. Highly pure polypeptide of interest is obtained by passing the cleaved polypeptide mixture over the avidin resin using the procedure described above.

The polypeptide for attachment is retained by the avidin monomer and the polypeptide of interest is not retained. The polypeptide of interest is collected in the column flowthrough.

Procedure XII. Plasmid Expression Vectors.

Two plasmids are obtained from D. Samols, Case Western Reserve University.

A. Plasmid ptac 1.3t. This plasmid contains the DNA sequence coding for the 123 amino acid sequence of the 1.3S polypeptide of transcarboxylase from *Propionibacterium shermanii* (SEQ ID NO:2). The DNA coding for the 1.3S polypeptide is cloned as a 431 base pair fragment into the polylinker region of the expression vector pKK223-3 as described by Murtif et al. (Proc Nat Acad Sci USA 82:5617-5621 (1985)).

B. Plasmid ptac1.3(1-125). The plasmid ptac 1.3(1-125) is described by Murtif and Samols, J Biol Chem 262:11813-11815 (1987). Like ptac1.3t, ptac1.3(1-125) also contains the 1.3S polypeptide but in addition has the sequence:

GAT CCA TAA CGC CTA AGC TT (SEQ ID NO:3)

at the 3' end of the 1.3S gene that encodes a BamHI restriction endonuclease site. This DNA additional sequence codes for the linking amino acid sequence asp-pro at the carboxyl terminus of the 1.3S polypeptide.

In order to illustrate the nature of this invention and the manner of practicing the same, the following examples are presented.

Example 1. Modification of ptac1.3(1-125) to increase hybrid polypeptide expression levels in E. coli.

In order to increase the expression level of hybrid polypeptides produced from chimeric genes inserted into ptac1.3(1-125) from approximately 0.1% of total soluble cellular protein to approximately 5.0% of total soluble cell protein, ptac1.3(1-125) was modified as follows. ptac1.3(1-125) was digested with the restriction enzymes XhoI and HindIII. The desired 131 base pair (bp) fragment was obtained by agarose gel electrophoresis.

The vector ptac1.3t was also digested with XhoI and HindIII using the conditions described above, and the 4.86 kilobase (kb) fragment was obtained by agarose gel purification. The plasmid ptac1.3dp (Figure 1) was obtained by ligation of the 131 bp fragment from ptac1.3(1-125) to the 4.86 kb fragment of ptac1.3t.

The ligated plasmid mixture was used to transform competent E. coli HB101. An E. coli clone harboring ptac1.3dp was identified by restriction enzyme digestion of plasmids isolated from selected ampicillin-resistant E. coli cells.

Example 2. Fusion of a polypeptide of interest to the C-terminus of a polypeptide for attachment with an acid cleavage site between the polypeptides

This example describes a hybrid polypeptide in which a synthetic b-endorphin polypeptide is fused to the carboxyl terminus of the 1.3S polypeptide. An asp-pro cleavage site is incorporated between the two polypeptides for cleavage and subsequent purification of b-endorphin away from the 1.3S polypeptide after affinity purification using avidin monomer resin.

The amino acid sequence of a modified b-endorphin polypeptide is shown in SEQ ID NO:6 and the corresponding nucleotide sequence coding for this amino acid sequence is shown in SEQ ID NO:7.

The synthetic oligonucleotides from which this synthetic gene was assembled are shown in SEQ ID NO:8 (RHcb1) and SEQ ID NO:9 (RHcb2). There are no internal methionine residues in the modified b-endorphin polypeptide.

5 The BamHI endonuclease cleavage recognition sequence GGATCC at nucleotide positions 12 through 17 at the 5' end of SEQ ID NO:6 allows the introduction of an ATG codon at the 5' terminus of the b-endorphin gene when this fragment is introduced into the BamHI site of ptac1.3dp (SEQ ID NO:5), thereby adding a methionine at the N-terminus of the polypeptide.

To maximize expression in *E. coli*, the amino acid sequence of authentic b-endorphin was reverse translated into a DNA sequence using preferred codon usages of highly expressed *E. coli* genes (DeBoer, H Chapter 8, Maximizing Gene Expression, W. Reznikoff and L. Gold, Eds.).

RHcb1 and RHcb2 were synthesized, annealed and filled by T7 DNA polymerase as described in Procedure VII.

The resulting double stranded DNA sequence (SEQ ID NO:5) was digested with BamHI, generating 105 bp fragment coding for the synthetic b-endorphin which was purified by agarose gel electrophoresis.

15 The plasmid vector pUC19 (Sambrook et al, Molecular Cloning, A Laboratory Manual, 2nd Edition, Vol 1, 1989, p. 1.13) was linearized with BamHI. The 5' terminus of the linearized plasmid was dephosphorylated prior to ligation by incubating the digest mixture with calf intestinal phosphatase using the protocol described by Sambrook et al (ibid, Vol.1, pp. 3.38-3.39), to minimize self-ligation of the vector.

20 Pure linear plasmid was recovered by agarose gel electrophoresis, and the 110 bp synthetic b-endorphin gene fragment was ligated to the pUC19 plasmid, and this ligated plasmid was used to transform competent *E. coli* HB101.

Following plasmid isolation from ampicillin-resistant clones, the recombinant *E. coli* cells harboring the correct plasmid were identified by restriction enzyme digestion. This recombinant plasmid containing the gene for the synthetic b-endorphin was designated pUC19endorB3.

25

Cloning of synthetic b-endorphin into ptac1.3dp.

30 The b-endorphin gene in pUC19endorB3 was fused to the 3' terminus of the 1.3S gene in ptac1.3dp as follows: a 105 bp b-endorphin gene fragment was generated by digestion of pUC19endorB3 with BamHI, and purified by agarose gel electrophoresis.

The vector ptac1.3dp, which contains two BamHI sites, was partially digested with BamHI. Plasmid DNA cut at only one BamHI site was purified by agarose gel electrophoresis.

The 105 bp b-endorphin gene was ligated into the BamHI site of the linearized ptac1.3dp plasmid and the ligation mixture was used to transform competent *E. coli* HB101.

35 The ligated plasmid containing the endorphin gene in the proper orientation was identified by restriction enzyme analysis of plasmids isolated from ampicillin-resistant transformed *E. coli* and was designated ptac1.3dp:endorB3 (Figure 2A). This plasmid codes for a hybrid fusion polypeptide consisting of the 1.3S polypeptide fused at its carboxyl terminus to an asp-pro cleavage sequence fused to a synthetic b-endorphin polypeptide containing a methionine residue at position 1.

40 A highly pure preparation of the synthetic b-endorphin was obtained by inoculation of L-broth containing 100 mg/l ampicillin with the *E. coli* host harboring ptac1.3dp:endorB3. A crude protein extract containing the 1.3S:b-endorphin hybrid polypeptide was obtained by following Procedure VII. Highly pure 1.3S:b-endorphin polypeptide was obtained by avidin monomer affinity chromatography described in Procedure VIII, using acetic acid to elute the purified hybrid polypeptide from the resin (Procedure IX A).

45 Cleavage of b-endorphin from the 1.3S polypeptide was accomplished by incubation in formic acid according to Part A of Procedure X, and highly pure b-endorphin was obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI. Following acid cleavage of an asp-pro linking sequence, a proline residue remains at the N-terminus of the cleaved b-endorphin polypeptide.

50 Clones containing the b-endorphin gene fragment inserted in the opposite orientation from ptac1.3dp:endorB3 yielded a 1.3S polypeptide fused to a novel 21 amino acid reverse endorphin peptide joined by the linking amino acid sequence asp-pro. The gene designated ptac1.3dp:revendorB3 that encodes this novel peptide is shown in Figure 2B. This polypeptide could be purified by avidin monomer chromatography (Procedure VIII), eluted in high yield and high purity from the column using acetic acid (Procedure IX A).

55 This example further demonstrates production of a hybrid polypeptide containing a polypeptide for binding to avidin as an efficacious method for obtaining polypeptides of interest in high yield and high purity.

Example 3. Fusion of a polypeptide of interest to a polypeptide for attachment, with a methionine cleavage site between the two polypeptides.

In this example, the synthetic b-endorphin polypeptide is fused at its N-terminus to a methionine residue, this methionine residue being positioned at the C-terminus of the 1.3S polypeptide, thus providing a single amino acid cleavage site for separation of b-endorphin from the 1.3S polypeptide following avidin monomer chromatography.

Cleavage with cyanogen bromide yields an unmodified N-terminus on b-endorphin, as the methionine is cleaved from the N-terminus of b-endorphin.

The vector ptac1.3dp is digested to completion with XhoI and HindIII, and the 131 bp fragment is purified by agarose gel electrophoresis. This 131 bp fragment is subjected to partial digestion with Sau3A, and the 110 bp fragment so generated is isolated and purified by agarose gel electrophoresis. This XhoI-Sau3A fragment is ligated to a double-stranded synthetic DNA fragment (SEQ ID NO:12) coding for b-endorphin. This b-endorphin has a methionine residue at its N-terminus. This fragment in SEQ ID NO:12 is assembled from synthetic oligonucleotides SEQ ID NO:10 and SEQ ID NO:11 as described in Procedure VI, and is digested to completion with Sau3A prior to ligation to the 110 bp XhoI-Sau3A fragment. The 220 bp product of this ligation is purified by agarose gel electrophoresis.

The vector ptac1.3dp is subjected to partial digestion with XhoI and BamHI, and the 4886 bp linear vector is purified by agarose gel electrophoresis. The ligated plasmid is used to transform competent E. coli CSH26, prepared according to Procedure IV. Plasmids are isolated from transformed ampicillin-resistant E. coli clones, and a plasmid containing the desired gene in the correct orientation is identified by restriction enzyme analysis and designated ptac1.3dp:met:endor (Figure 2C).

A highly pure preparation of the synthetic b-endorphin is obtained by inoculation of L-broth containing 100 mg/l ampicillin with the E. coli host harboring ptac1.3:met:endor. A crude protein extract containing the 1.3S:b-endorphin hybrid polypeptide is obtained by following Procedure VII.

Highly pure 1.3S:b-endorphin polypeptide is obtained by avidin monomer affinity chromatography described in Procedure VIII, using elution with acetic acid to elute the purified hybrid polypeptide from the resin (procedure VIIIA). Cleavage of b-endorphin from the 1.3S polypeptide was accomplished by incubation in cyanogen bromide according to Part B of Procedure X, and highly pure b-endorphin is obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI.

Example 4. Fusion of a polypeptide of interest directly to the carboxyl terminus of a polypeptide for attachment, with no linking amino acid sequence being present in the hybrid polypeptide.

In this example, a synthetic b-endorphin polypeptide is fused directly to the C-terminus of the 1.3S polypeptide. Avidin monomer chromatography is used to obtain highly pure b-endorphin in the form of a hybrid fusion polypeptide.

The construct ptac1.3t (Procedure XII) is digested to completion with XhoI and HindIII, and the smaller 131 bp fragment is purified by agarose gel electrophoresis. This 131 bp fragment is subjected to partial digestion with Sau3A I, and the 110 bp fragment is agarose-gel purified. The 110 bp fragment is ligated to the double stranded DNA fragment SEQ ID NO:15.

SEQ ID NO:15 encodes a synthetic b-endorphin gene with no DNA coding for a linking amino acid or amino acid sequence at its 3' terminus.

SEQ ID NO:15 is assembled from synthetic oligonucleotides SEQ ID NO:13 and SEQ ID NO:14 using Procedure VI.

Prior to ligation to the 110 bp fragment, SEQ ID NO:15 is digested to completion with Sau3A I. The 217 bp ligation product is purified by agarose gel electrophoresis.

Vector ptac1.3dp is linearized by partial digestion with XhoI and BamHI, and the 4886 bp fragment is also purified using agarose gel electrophoresis.

The ligated plasmid is used to transform competent E. coli CSH26.

Recombinant plasmids are isolated from ampicillin-resistant transformants, and a clone containing the desired gene in the correct orientation is identified by restriction enzyme analysis and designated ptac1.3:endor (Figure 2D).

A highly pure preparation of the synthetic b-endorphin is obtained by inoculation of L-broth containing 100 mg/l ampicillin with the E. coli host harboring ptac1.3:endor.

A crude protein extract containing the the 1.3S:b-endorphin hybrid polypeptide is obtained by following Procedure VII.

Highly pure 1.3S:b-endorphin polypeptide is obtained by avidin monomer affinity chromatography as de-

scribed in Procedure VIII, using elution with acetic acid to elute the purified hybrid polypeptide from the resin (procedure VIIIA).

5 **Example 5. Fusion of two polypeptides of interest to the C-terminus of a single polypeptide for attachment in the order, polypeptide for attachment:polypeptide of interest:polypeptide of interest.**

In this example, two b-endorphin polypeptides in tandem are fused to the C-terminus of one 1.3S polypeptide, with the linking amino acids asp-pro-met separating the first b-endorphin polypeptide from the C-terminus of the 1.3S polypeptide and another sequence of asp-pro-met separating the first b-endorphin from the second
10 b-endorphin polypeptide. Such a fusion doubles the yield of the polypeptide of interest, at the same time providing a means for purification of that polypeptide by avidin affinity chromatography.

The plasmid ptac1.3dp (Figure 1) is digested to completion with XhoI and BamHI. The resulting 118 bp fragment that encodes for the sequence beginning at amino acids 8 to the asp-pro site generated by the BamHI site at the 3' terminus of the 1.3S polypeptide as found in SEQ ID NO:5. This fragment is purified by agarose
15 gel electrophoresis.

Two tandem b-endorphin polypeptides are generated by the creation of the double-stranded DNA fragments SEQ ID NO:18, assembled from oligonucleotides SEQ ID NO:16 and SEQ ID NO:17; and SEQ ID NO:21, assembled from SEQ ID NO:19 and SEQ ID NO:20.

20 SEQ ID NO:18 and SEQ ID NO:21 were assembled from their respective oligonucleotides using the synthesis and strand assembly strategy described in Procedure VI.

SEQ ID NO:18 and SEQ ID NO:21 are digested with BamHI and ligated.

This fragment codes for two b-endorphin polypeptides in tandem, separated from each other by an asp-pro cleavage sequence. The dimeric product of ligation is purified by agarose gel electrophoresis, and this fragment is ligated to the 118 bp XhoI-BamHI 1.3S partial coding sequence obtained from ptac1.3dp. This ligation
25 product is purified by agarose gel electrophoresis, and is ligated to a 4886 bp fragment generated by partial digestion of ptac1.3dp linearized by partial digestion with XhoI and BamHI.

Plasmid DNA is isolated from transformed E. coli HB101. Plasmids containing the correct chimeric gene orientation are confirmed by restriction endonuclease mapping and designated ptac1.3: endor:endor (Figure 3A).

30 A highly pure preparation of synthetic b-endorphin is obtained by inoculation of L-broth containing 100 mg/l ampicillin with the E. coli host harboring ptac1.3:endor:endor.

A crude protein extract containing the 1.3S:b-endorphin: b-endorphin hybrid polypeptide is obtained by following Procedure VII. Highly pure 1.3S:b-endorphin: b-endorphin polypeptide is obtained by avidin monomer affinity chromatography described in Procedure VIII, using acetic acid to elute the purified hybrid polypeptide
35 from the resin (procedure VIIIA).

Cleavage of both b-endorphin polypeptides from the 1.3S polypeptide in a single step is accomplished by incubation in formic acid according to Part A of Procedure X, and highly pure b-endorphin is obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI.

40 **Example 6. Fusion of one polypeptide of interest to the N-terminus of a polypeptide for attachment and fusion of a second polypeptide of interest to the C-terminus of the same polypeptide for attachment.**

In this example, the maltose binding protein (Guan, C. et al., Gene 67-21-30 (1987) and Maina, et al., Gene 74:365-373 (1988)) was fused to the N-terminus of the 1.3S polypeptide and synthetic b-endorphin was fused
45 to the C-terminus of the same 1.3S polypeptide, thus creating a hybrid polypeptide consisting of two different noncontiguous polypeptides of interest.

The construct ptac1.3dp:endorB3 (Figure 2A) was digested with Sall and HindIII. A 438 bp fragment created was purified by agarose gel electrophoresis. This fragment encodes amino acids 19 to 123 of the 1.3S polypeptide, the asp-pro-met linking amino acids, and the 31 amino acid b-endorphin polypeptide.

50 The vector pMAL-c (obtained from New England Biolabs) was linearized by digestion with Sall and HindIII. This vector contains the maltose binding protein under the regulation of the tac promoter (Guan, C. et al., Gene 67-21-30 (1987) and Maina, et al., Gene 74:365-373 (1988)). The linearized vector and the 438 bp 1.3S-b-endorphin fragment were ligated, and the ligation mix was used to transform competent E. coli CHS26. Plasmid DNA was isolated, and plasmids containing the correct chimeric gene orientation were confirmed by restriction
55 endonuclease mapping. The resulting clone was designated ptac:malB:1.3:endorB3 (Figure 3B).

A highly pure preparation of the hybrid maltose binding protein-synthetic b-endorphin polypeptide is obtained by inoculation of L-broth containing 100 mg/l ampicillin with the E. coli host harboring ptac:malB: 1.3:endorB3.

A crude protein extract containing the hybrid polypeptide is obtained by following Procedure VII.

Highly pure maltose binding protein:1.3S:b- endorphin hybrid polypeptide is obtained by avidin monomer affinity chromatography described in Procedure VIII, using biotin to elute the purified hybrid polypeptide from the resin (Procedure VIIIB).

5 Biotin is removed from the polypeptide suspension by dialysis against three changes of 100 mM ammonium carbonate buffer, pH 7.2, followed by freeze-drying of the sample.

Cleavage of b-endorphin from the 1.3S polypeptide is accomplished by incubation in cyanogen bromide according to Part B of Procedure X, and highly pure b endorphin is obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI.

10 The maltose-binding protein-1.3S hybrid polypeptide was recovered in highly pure form by repeating the biotin elution procedure described in Procedure VIIIB.

Example 7. Fusion of two polypeptides of interest in tandem to the N-terminus of a polypeptide for attachment, with an amino acid cleavage sequence separating the first polypeptide of interest from the polypeptide for attachment, and a second amino acid cleavage sequence separating the first polypeptide of interest from the second polypeptide of interest.

The maltose binding protein and b-endorphin are fused in tandem to the amino-terminus of the 1.3S polypeptide. An amino acid cleavage site separates the maltose binding protein and b-endorphin, and another amino acid cleavage site separates b-endorphin and the 1.3S polypeptide.

The plasmid ptac1.3dp is digested with HincII and HindIII. The 332 bp fragment is purified by agarose gel electrophoresis. Linkers encoding a BamHI recognition sequence (CGGATCCG) are ligated to this fragment, and the fragment is digested with BamHI to generate a BamHI site at the 5' terminus of the fragment. The 338 bp BamHI -HindIII fragment so generated is purified by agarose gel electrophoresis .

25 The DNA fragment SEQ ID NO:15 is digested with BamHI , and is ligated to the 338 bp BamHI-HindIII modified fragment from ptac1.3dp. The desired 444 bp fragment is purified by agarose gel electrophoresis. The vector pMAL-c (New England Biolabs, Beverly, MA) is digested to a 6.1 kb fragment with HindIII and BamHI , and the large fragment is purified by agarose gel electrophoresis. The 444 bp fragment and the 6.1 kb fragment are ligated, and the ligation mix is used to transform competent E. coli CSH26.

30 Plasmid DNA is isolated, and plasmids containings the chimeric gene in the correct orientation are confirmed by restriction endonuclease mapping, and are designated ptac:malC:endorB3: 1.3dp (Fig. 3C). The correct recombinant plasmid codes for a fusion protein composed of a 42,000 MW maltose binding protein fused by an asp-pro-met linker to b-endorphin joined by an asp-pro linker to amino acids 19-123 of the 1.3S polypeptide having an asp-pro carboxyl terminus.

35 Highly pure maltose binding protein:b- endorphin: 1.3S hybrid polypeptide is obtained by avidin monomer affinity chromatography described in Procedure VIII, using biotin to elute the purified hybrid polypeptide from the resin (Procedure VIIIB). Biotin is removed from the polypeptide suspension by dialysis against three changes of 100 mM ammonium carbonate buffer, pH 7.2, followed by freeze-drying of the sample.

40 Cleavage of b-endorphin from the 1.3S polypeptide is accomplished by incubation in formic acid according to Part A of Procedure X, and highly pure b-endorphin is obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI. The maltose binding protein is recovered in highly pure form by elution of the maltose binding protein:1.3S hybrid polypeptide from the avidin monomer resin with biotin using Part B of Procedure IX.

45 The hybrid polypeptide is purified away from the biotin by dialysis against three changes of 100 mM ammonium carbonate buffer, pH 7.2, followed by freeze-drying of the sample. The sample is freeze-dried, and reconstituted in cyanogen bromide according to Part B of Procedure X. The maltose-binding protein is recovered in highly pure form by repeating the avidin monomer chromatography process detailed in Procedure XI.

Example 8. Fusion of two polypeptides of interest to the C-termini of two polypeptides for attachment within the same hybrid polypeptide.

55 In this example, two noncontiguous b-endorphin polypeptides are fused to two noncontiguous 1.3S polypeptides within one hybrid polypeptide with a cleavage amino acid sequence between each 1.3S polypeptide and the b-endorphin to which it is directly linked, producing the fusion hybrid polypeptide 1.3S:asp-pro-met: b-endorphin:asp-pro:1.3S:asp-pro-met:b-endorphin.

The vector 1.3dp:endorB3 (Figure 2A) is digested with HincII and HindIII and the 437 bp fragment is purified by agarose gel electrophoresis. Synthetic DNA linkers encoding a BamHI recognition sequence, CGGATCCG (New England Biolabs, Inc. Beverly, MA) are ligated to this 437 bp fragment.

Following ligation, this DNA is subjected to BamHI digestion to generate BamHI cohesive termini, and the 443 bp BamHI-HindIII fragment is purified by agarose gel electrophoresis. The ptac 1.3:endor:endor (Figure 3A) is partially digested with BamHI and HindIII to yield a linear DNA of 5.1 kb digested at a single BamHI site and also at a single HindIII site. This 5.1 kb fragment is purified by agarose gel electrophoresis, then is ligated to the 443 bp fragment. The ligation mix is used to transform competent E. coli HB101.

Plasmid DNA is isolated from ampicillin-resistant transformed E. coli, and plasmids containing the correct chimeric gene orientation are confirmed by restriction endonuclease mapping.

The recombinant plasmid obtained by this procedure is designated ptac1.3:endor: 1.3:endor (Figure 4) and encodes a hybrid polypeptide which permits the isolation of two molecules of b-endorphin for every hybrid polypeptide purified, which may double the yield of polypeptide from a single fermentation.

A highly pure preparation of synthetic b-endorphin is obtained by inoculation of L-broth containing 100 mg/l ampicillin with the E. coli host harboring ptac1.3: endor:1.3:endor. A crude protein extract containing the 1.3S:b-endorphin:1.3S:b-endorphin hybrid polypeptide is obtained by following Procedure VII.

Highly pure 1.3S: b-endorphin: 1.3S :b-endorphin polypeptide is obtained by avidin monomer affinity chromatography described in Procedure VIII, using acetic acid to elute the purified hybrid polypeptide from the resin (procedure VIIIA).

Cleavage of both b-endorphin polypeptides from both 1.3S polypeptides in a single step is accomplished by incubation in formic acid according to Part A of Procedure X, and highly pure b-endorphin is obtained by avidin monomer affinity chromatography of the cleavage mixture by repeating Procedure XI.

SEQ ID NO: 2

5 SEQUENCE TYPE: Peptide
 SEQUENCE LENGTH: 43
 MOLECULE TYPE: Peptide
 10 ORIGINAL SOURCE ORGANISM: Bacterium
 SOURCE NAME: *Propionibacterium shermanii*
 PROPERTIES: Biotin-binding recognition sequence
 15
 Pro Ala Pro Leu Ala Gly Thr Val Ser Lys Ile Leu Val Lys Glu Gly
 1 5 10 15
 20 Asp Thr Val Lys Ala Gly Gln Thr Val Leu Val Leu Glu Ala Met Lys
 20 25 30
 Met Glx Thr Glu Ile Asn Ala Pro Thr Asp Gly
 25 35 40

SEQ ID NO: 3

30
 SEQUENCE TYPE: Nucleotide
 SEQUENCE LENGTH: 20
 35 STRANDEDNESS: Double-stranded
 TOPOLOGY: Linear
 MOLECULE TYPE: Genomic DNA
 40 PROPERTIES: Termination fragment for a BamHI cleavage site

45 GATCCATAAC GCCTAAGCTT 20
 50
 55

SEQ ID NO:4

5

SEQUENCE TYPE: Nucleotide

10

SEQUENCE LENGTH: 372

15

STRANDEDNESS: Double-stranded

20

TOPOLOGY: Linear

MOLECULE TYPE: Genomic DNA

25

ORIGINAL SOURCE ORGANISM: Bacterium

SOURCE NAME: *Propionibacterium shermanii*

30

PROPERTIES: Gene coding for 1.3S polypeptide

35

ATGAAACTGA AGGTAACAGT CAACGGCACT GCGTATGACG	40
TTGACGTTGA CGTCGACAAG TCACACGAAA ACCCGATGGG	80
CACCATCCTG TTCGGCGGCG GCACCGGCGG CGCGCCGGCA	120
CCGCGCGCAG CAGGTGGCGC AGGCGCCGGT AAGGCCGGAG	160
AGGGCGAGAT TCCCGCTCCG CTGGCCGGCA CCGTCTCAA	200
GATCCTCGTG AAGGAGGGTG ACACGGTCAA GGCTGGTCAG	240
ACCGTGCTCG TTCTCGAGGC CATGAAGATG GAGACCGAGA	280
TCAACGCTCC CACCGACGGC AAGGTCGAGA AGGTCCTTGT	320
CAAGGAGCGT GACGCCGTGC AGGGCGGTCA GGGTCTCATC	360
AAGATCGGCT GA	372

50

55

SEQ ID NO:5

5 SEQUENCE TYPE: Nucleotide
 SEQUENCE LENGTH: 390
 STRANDEDNESS: Double-stranded
 10 TOPOLOGY: Linear
 MOLECULE TYPE: Genomic DNA
 PROPERTIES: Oligonucleotide ptac1.3dp

15
 ATGAAACTGA AGGTAACAGT CAACGGCACT GCGTATGACG 40
 TTGACGTTGA CGTCGACAAG TCACACGAAA ACCCGATGGG 80
 20 CACCATCCTG TTCGGCGGCG GCACCGGCGG CGCGCCGGCA 120
 CCGCGCGCAG CAGGTGGCGC AGGCGCCGGT AAGGCCGGAG 160
 AGGGCGAGAT TCCCGCTCCG CTGGCCGGCA CCGTCTCCAA 200
 25 GATCCTCGTG AAGGAGGGTG ACACGGTCAA GGCTGGTCAG 240
 ACCGTGCTCG TTCTCGAGGC CATGAAGATG GAGACCGAGA 280
 TCAACGCTCC CACCGACGGC AAGGTCGAGA AGGTCCTTGT 320
 30 CAAGGAGCGT GACGCCGTGC AGGGCGGTCA GGGTCTCATC 360
 AAGATCGGCT GATCCATAAC GCCTAAGCTT 390

35 SEQ ID NO:6

SEQUENCE TYPE: Peptide
 SEQUENCE LENGTH: 31
 40 MOLECULE TYPE: Peptide
 PROPERTIES: Modified b-endorphin polypeptide

45 Tyr Gly Gly Phe Leu Thr Ser Glu Lys Ser Gln Thr Pro Leu Val Thr
 1 5 10 15
 50 Leu Phe Lys Asn Ala Ile Ile Lys Asn Ala Tyr Lys Lys Gly Glu
 20 25 30

55

SEQ ID NO:7

5 SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 135
STRANDEDNESS: Double-stranded
10 TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
PROPERTIES: Gene coding for modified b-endorphin polypeptide

15

AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT 40
20 CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTCAAAAA 80
CGCTATCATC AAAAACGCAT ACAAAAAAGG CGAATAAGGA 120
TCCGAATTCG AGCTC 135

25

SEQ ID NO: 8

30

SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 75
35 STRANDEDNESS: Double-stranded
TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
40 PROPERTIES: Oligonucleotide RHcbe1

45

AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT 40
CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTTC 75

50

55

SEQ ID NO: 9

5 **SEQUENCE TYPE: Nucleotide**
 SEQUENCE LENGTH: 75
 STRANDEDNESS: Double-stranded
10 **TOPOLOGY: Linear**
 MOLECULE TYPE: Genomic DNA
 PROPERTIES: Oligonucleotide RHcbe2

15

GAGCTCGAAT TCGGATCCTT ATTCGCCTTT TTTGTATGCG 40
20 **TTTTTGATGA TAGCGTTTTT GAACAGAGTA ACCAG 75**

25

SEQ ID NO. 10

SEQUENCE TYPE: Nucleotide
30 **SEQUENCE LENGTH: 75**
 STRANDEDNESS: Double-stranded
 TOPOLOGY: Linear
35 **MOLECULE TYPE: Genomic DNA**
 PROPERTIES: Synthetic oligonucleotide

40

AAGCTTCTAG AGATCGGCAT GTACGGTGGT TTCCTGACCT 40
45 **CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTTC 75**

50

55

SEQ ID NO:11

5 SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 75
STRANDEDNESS: Double-stranded
10 TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
PROPERTIES: Synthetic oligonucleotide

15

GAGCTCGAAT TCGGATCCTT ATTCGCCTTT TTTGTATGCG 40
20 TTTTGTATGA TAGCGTTTTT GAACAGAGTA ACCAG 75

25 SEQ ID NO:12

SEQUENCE TYPE: Nucleotide
30 SEQUENCE LENGTH: 135
STRANDEDNESS: Double-stranded
TOPOLOGY: Linear
35 MOLECULE TYPE: Genomic DNA

40 AAGCTTCTAG AGATCGGCAT GTACGGTGGT TTCCTGACCT 40
CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTCAAAAA 80
CGCTATCATC AAAAACGCAT ACAAAAAAGG CGAATAAGGA 120
45 TCCGAATTCG AGCTC 135

50

55

SEQ IN NO:13

5 SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 75
STRANDEDNESS: Double-stranded
10 TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
PROPERTIES: Synthetic oligonucleotide

15

AAGCTTCTAG AGATCGGCTA CGGTGGTTTC CTGACCTCCG 40
20 AAAAATCTCA GACCCCGCTG GTTACTCTGT TCAAA 75

25

SEQ ID NO:14

SEQUENCE TYPE: Nucleotide
30 SEQUENCE LENGTH: 72
STRANDEDNESS: Double-stranded
TOPOLOGY: Linear
35 MOLECULE TYPE: Genomic DNA
PROPERTIES: Synthetic oligonucleotide

40

GAGCTCGAAT TCGGATCCTT ATTCGCCTTT TTTGTATGCG 40
45 TTTTGATGA TAGCGTTTTT GAACAGAGTA AC 72

50

55

SEQ ID NO:15

5 SEQUENCE TYPE: Nucleotide
 SEQUENCE LENGTH: 132
 STRANDEDNESS: Double-stranded
 10 TOPOLOGY: Linear
 MOLECULE TYPE: Genomic DNA

15

	AAGCTTCTAG AGATCGGCTA CGGTGGTTTC CTGACCTCCG	40
	AAAAATCTCA GACCCCGCTG GTTACTCTGT TCAAAAACGC	80
20	TATCATCAAA AACGCATACA AAAAAGGCGA ATAAGGATCC	120
	GAATTCGAGC TC	132

25

SEQ ID NO:16

30

SEQUENCE TYPE: Nucleotide
 SEQUENCE LENGTH: 75
 STRANDEDNESS: Double-stranded
 35 TOPOLOGY: Linear
 MOLECULE TYPE: Genomic DNA
 40 PROPERTIES: Synthetic oligonucleotide

45

	AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT	40
	CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTT	75

50

55

SEQ ID NO:17

5 SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 71
STRANDEDNESS: Double-stranded
10 TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
PROPERTIES: Synthetic oligonucleotide

15

GAGCTCGAAT TCGGATCCTC GCCTTTTTTG TATGCGTTTT 40
TGATGATAGC GTTTTTGAAC AGAGTAACCA G 71

20

SEQ ID NO:18

25

SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 130
30 STRANDEDNESS: Double-stranded
TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA

35

40 AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT 40
CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTCAAAAA 80
CGCTATCATC AAAAACGCAT ACAAAAAAAGG CGAGGATCCG 120
45 AATTCGAGCTC 130

50

55

SEQ ID NO:19

5 **SEQUENCE TYPE: Nucleotide**
 SEQUENCE LENGTH: 75
 STRANDEDNESS: Double-stranded
10 **TOPOLOGY: Linear**
 MOLECULE TYPE: Genomic DNA
 PROPERTIES: Synthetic oligonucleotide
15

AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT 40
20 **CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTC 75**

25 **SEQ ID NO:20**

SEQUENCE TYPE: Nucleotide
30 **SEQUENCE LENGTH: 75**
 STRANDEDNESS: Double-stranded
 TOPOLOGY: Linear
35 **MOLECULE TYPE: Genomic DNA**
 PROPERTIES: Synthetic oligonucleotide

40
 GAGCTCGAAT TCGGATCCTT ATTCGCCTTT TTTGTATGCG 40
45 **TTTTTGATGA TAGCGTTTTT GAACAGAGTA ACCAG 75**

50

55

SEQUENCE LISTING

5 SEQ. ID NO:1

SEQUENCE TYPE: Peptide

10 SEQUENCE LENGTH: 123

MOLECULE TYPE: Protein

ORIGINAL SOURCE ORGANISM: Bacterium

15 SOURCE NAME: *Propionibacterium shermanii*

FEATURES: From 58 to 100 - biotin-binding recognition sequence

PROPERTIES: 1.3S biotin-binding protein

20

Met Lys Leu Lys Val Thr Val Asn Gly Thr Ala Tyr Asp Val Asp Val
 25 1 5 10 15
 Asp Val Asp Lys Ser His Glu Asn Pro Met Gly Thr Ile Leu Phe Gly
 20 25 30
 30 Gly Gly Thr Gly Gly Ala Pro Ala Pro Arg Ala Ala Gly Gly Ala Gly
 35 40 45
 Ala Gly Lys Ala Gly Glu Gly Glu Ile Pro Ala Pro Leu Ala Gly Thr
 35 50 55 60
 Val Ser Lys Ile Leu Val Lys Glu Gly Asp Thr Val Lys Ala Gly Gln
 40 65 70 75 80
 Thr Val Leu Val Leu Glu Ala Met Lys Met Glx Thr Glu Ile Asn Ala
 85 90 95
 45 Pro Thr Asp Gly Lys Val Glu Lys Val Leu Val Lys Glu Arg Asp Ala
 100 105 110
 Val Gln Gly Gly Gln Gly Leu Ile Lys Ile Gly
 50 115 120

55

SEQ ID NO:21

5

SEQUENCE TYPE: Nucleotide

SEQUENCE LENGTH: 135

10

STRANDEDNESS: Double-stranded

TOPOLOGY: Linear

MOLECULE TYPE: Genomic DNA

15

AAGCTTCTAG AGGATCCTAT GTACGGTGGT TTCCTGACCT 40
CCGAAAAATC TCAGACCCCG CTGGTTACTC TGTTCAAAAA 80
CGCTATCATC AAAAACGCAT AAAAAAAGG CGAATAAGGA 120
TCCGAATTCG AGCTC 135

25

30

Claims

1. A recombinant hybrid polypeptide comprising a polypeptide of interest fused to an avidin-binding polypeptide containing a biotin attachment domain, characterised in that the polypeptide of interest is fused to the C terminus of the avidin-binding polypeptide.
2. A recombinant hybrid polypeptide according to claim 1 wherein biotin is attached to the avidin-binding polypeptide.
3. A recombinant hybrid polypeptide according to claim 1 or claim 2 wherein the polypeptide includes a cleavage site for cleaving the polypeptide of interest from the avidin-binding polypeptide.
4. A recombinant hybrid polypeptide according to any one of claims 1 to 3 wherein the avidin binding polypeptide is, or is part of, a 1.3S polypeptide.
5. A recombinant hybrid polypeptide according to claim 4 wherein the 1.3S polypeptide is from Propionibacterium.
6. A recombinant hybrid polypeptide according to any one of claims 1 to 5 wherein the biotin attachment domain of the avidin-binding polypeptide comprises at least one of the sequence Pro Ala Pro Leu Ala Gly Thr Val Ser Lys Ile Leu Val Lys Glu Gly Asp Thr Val Lys Ala Gly Gln Thr Val Leu Val Leu Glu Ala Met Lys Met Glu Thr Glu Ile Asn Ala Pro Thr Asp Gly.
7. A recombinant hybrid polypeptide according to any one of claims 1 to 6 wherein the avidin-binding polypeptide comprises a plurality of non-contiguous and/or contiguous avidin-binding polypeptides, which may be the same or different.
8. A recombinant hybrid polypeptide according to any one of claims 1 to 6 wherein the polypeptide of interest

- comprises a plurality of non-contiguous and/or contiguous polypeptides of interest, which may be the same or different.
- 5 9. A recombinant hybrid polypeptide according to any one of claims 1 to 8 wherein the polypeptide of interest is an enzyme, is an antigen useful for vaccine production, or a diagnostic reagent.
 - 10 10. A recombinant hybrid polypeptide according to any one of claims 1 to 9 wherein the polypeptide of interest has antitumor activity or has an amino acid sequence for recognition of antigens.
 - 11 11. A nucleic acid sequence coding for a hybrid polypeptide as defined in any one of claims 1 to 10 wherein the nucleic acid sequence comprises a nucleic acid sequence coding for an avidin-binding polypeptide upstream of a nucleic acid sequence coding for a polypeptide of interest.
 - 12 12. A nucleic acid sequence according to claim 11 wherein the nucleic acid sequence is a DNA sequence.
 - 15 13. A nucleic acid sequence according to claim 12 wherein the DNA sequence contains in a 5' to 3' direction on the coding strand a gene comprising a 5' promoter region, the DNA sequence coding for the avidin-binding polypeptide and the DNA sequence coding for the polypeptide of interest.
 - 20 14. A nucleic acid sequence according to claim 12 or claim 13 wherein the DNA sequence is, or is part of, an expression vector or a plasmid.
 - 25 15. A process for the production of a hybrid polypeptide as defined in any one of claims 1 to 10 comprising constructing a plasmid containing a nucleic acid sequence as defined in any one of claims 11 to 13, transforming the plasmid into a procaryotic or eucaryotic host cell expression system, expressing the system, contacting the hybrid polypeptide resulting from the expression system with avidin, and harvesting the resulting avidin-bound hybrid polypeptide.
 - 30 16. A process according to claim 15 wherein the expression system is either E. Coli or insect cells.
 - 31 17. A process for the isolation of a hybrid polypeptide as defined in any one of claims 1 to 10 comprising contacting the hybrid polypeptide with avidin.
 - 32 18. A process according to any one of claims 15 to 17 wherein the avidin is monomeric avidin, tetrameric avidin or streptavidin.
 - 35 19. A process according to any one of claims 15 to 18 wherein the polypeptide of interest is cleaved from the isolated hybrid polypeptide.
 - 40 20. A process according to any one of claims 15 to 19 wherein the hybrid polypeptide is isolated using avidin covalently bound to a chemically inert, solid, water and solvent insoluble substrate through a chemically stable non-hydrolyzable linking group, preferably the hybrid polypeptide is isolated using avidin monomer affinity chromatography.
 - 41 21. A first kit comprising a hybrid polypeptide as defined in any one of claims 1 to 10 and avidin.
 - 45 22. A second kit comprising a nucleic acid sequence as defined in any one of claims 11 to 14 and avidin.
 - 46 23. A third kit comprising a nucleic acid sequence which codes for an avidin-binding polypeptide containing a biotin attachment domain and which is fusable to a nucleic acid sequence coding for a polypeptide of interest in order to form a hybrid nucleic acid sequence as defined in any one of claims 11 to 14 and avidin.
 - 50 24. A kit according to claim 23 wherein the kit comprises means to fuse the nucleic acid sequence coding for the avidin-binding polypeptide to the nucleic acid sequence coding for the polypeptide of interest in order to form the nucleic acid sequence as defined in any one of claims 11 to 14.
 - 55 25. A kit according to any one of claims 21 to 24 wherein the kit comprises means to cleave the polypeptide of interest from the avidin-binding polypeptide.

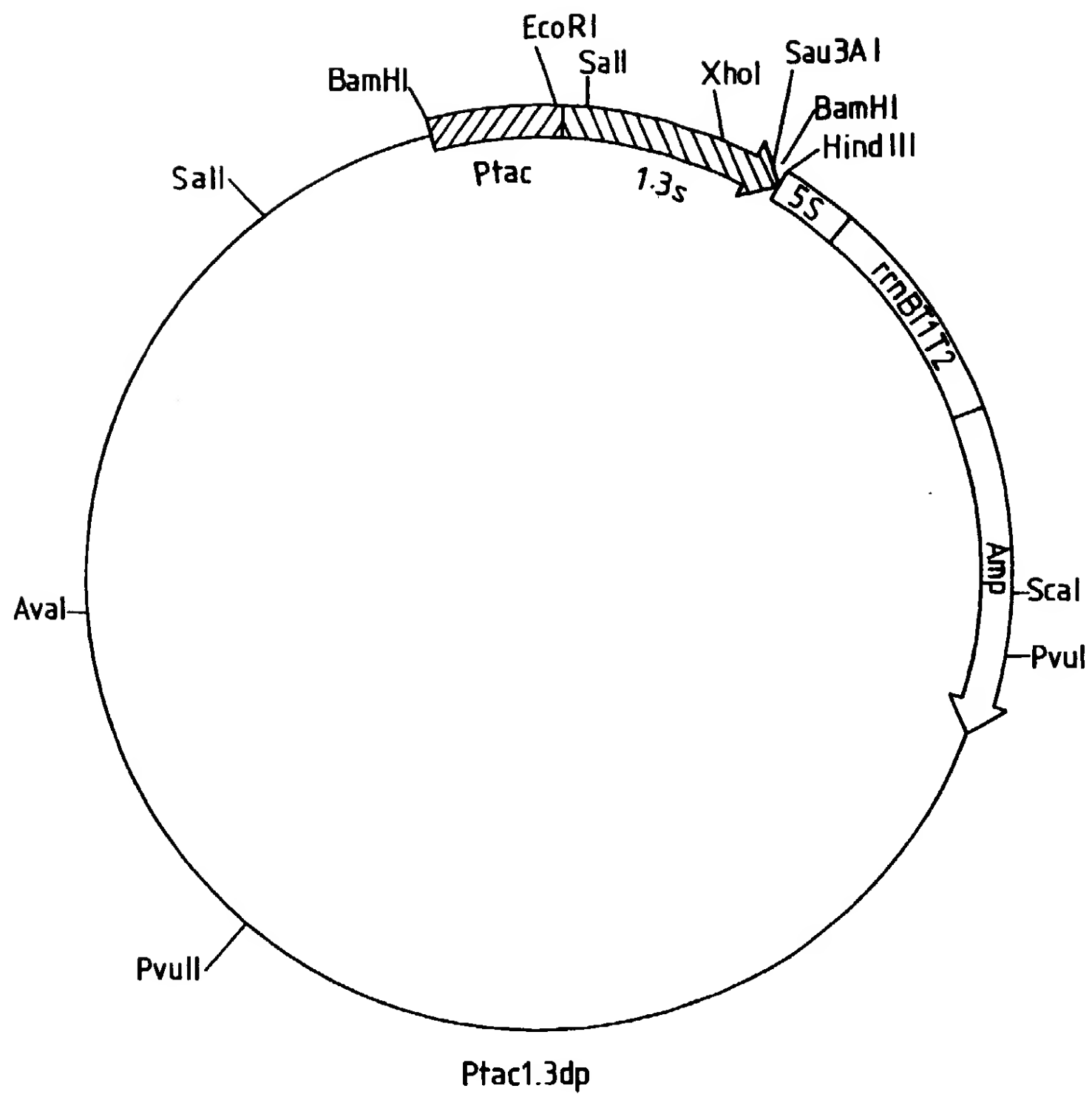


Fig.1.

Fig.2.

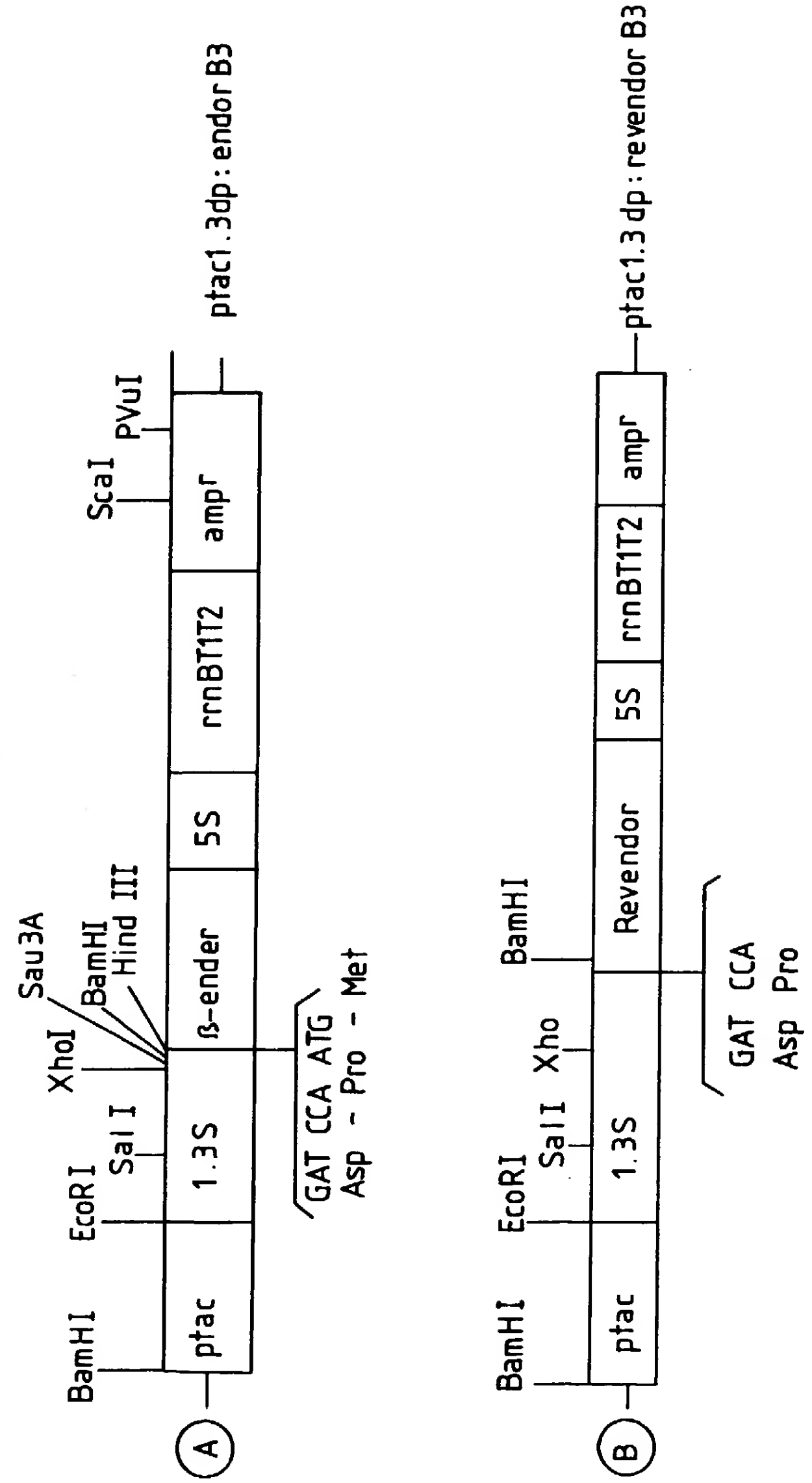


Fig.2. (Continued)

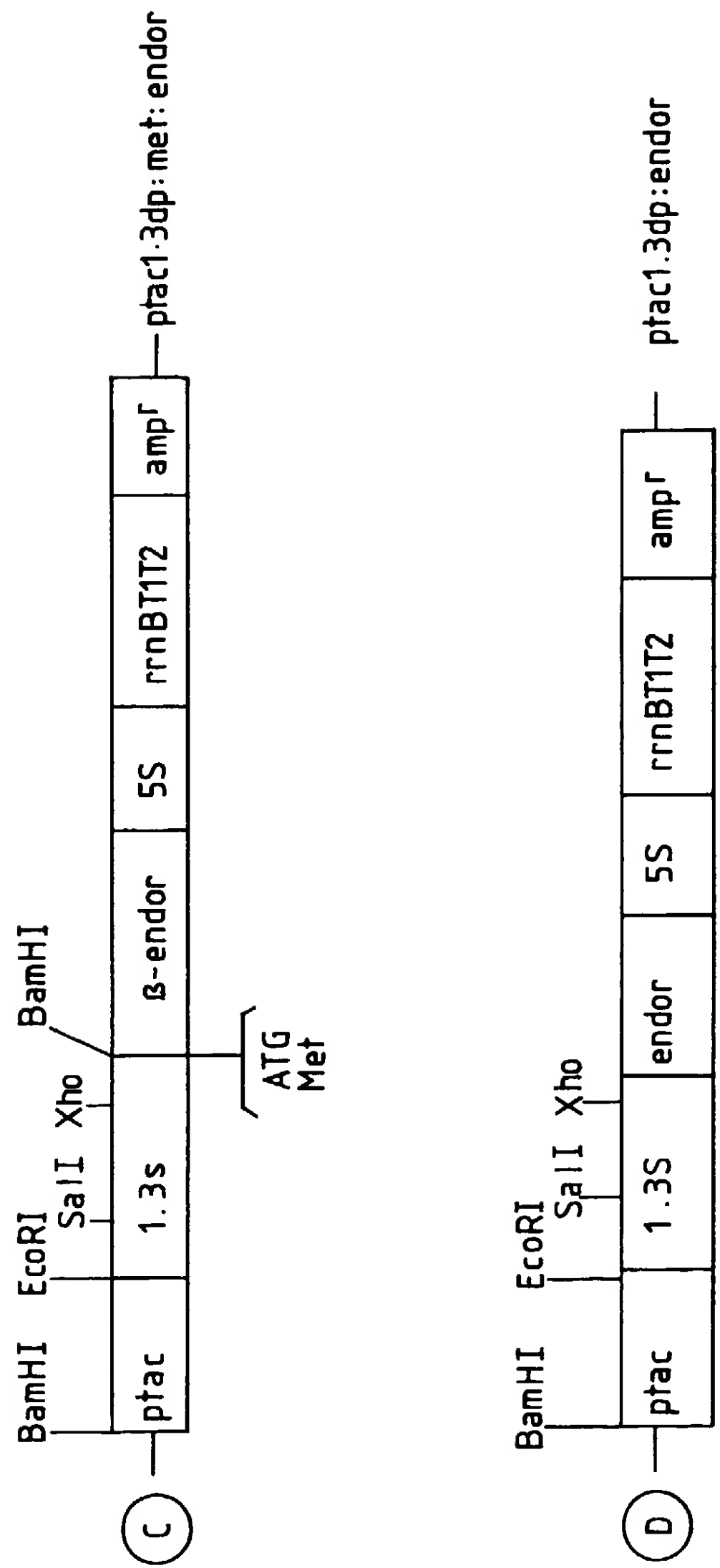


Fig. 3.

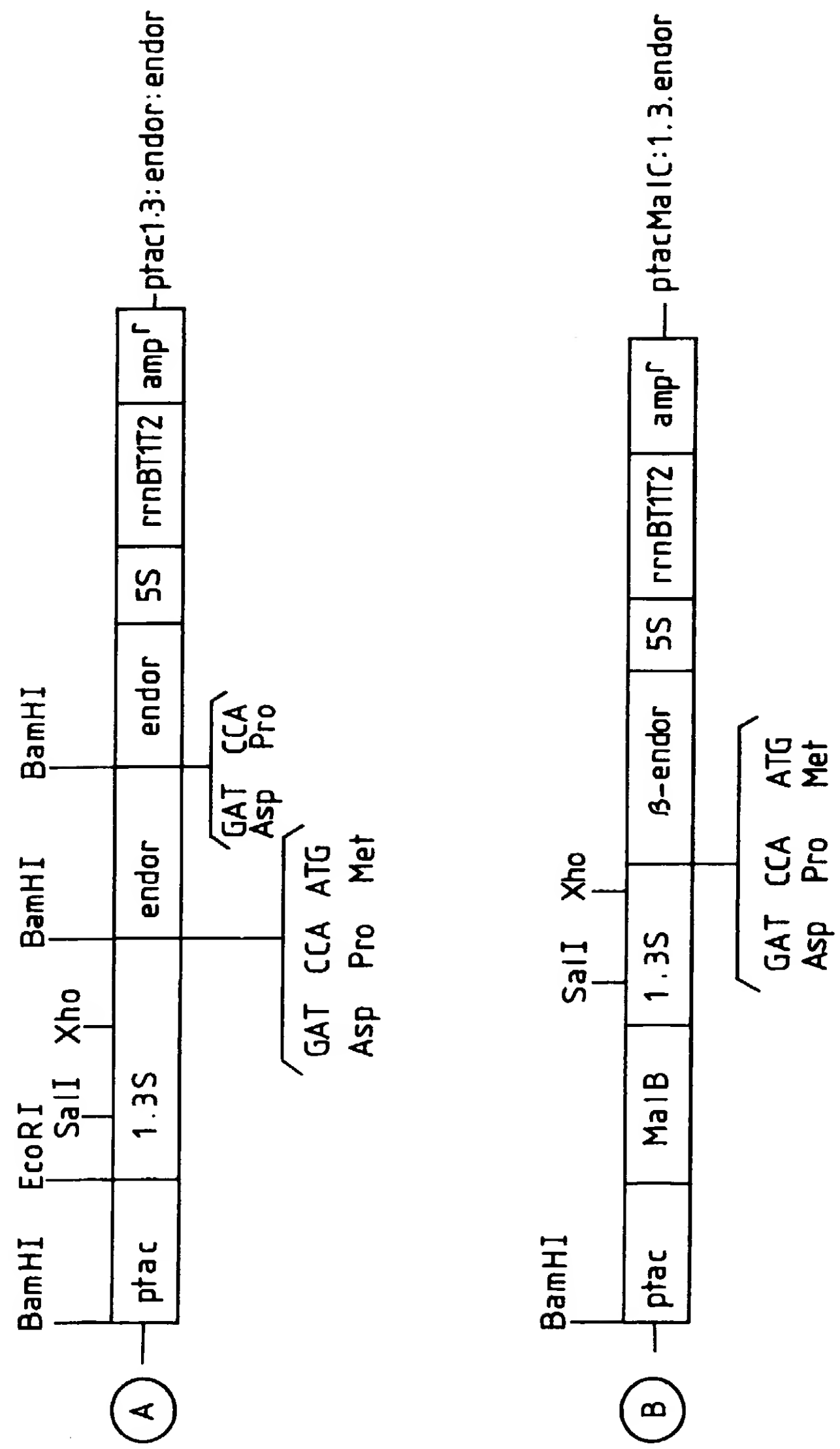


Fig.3. (Continued)

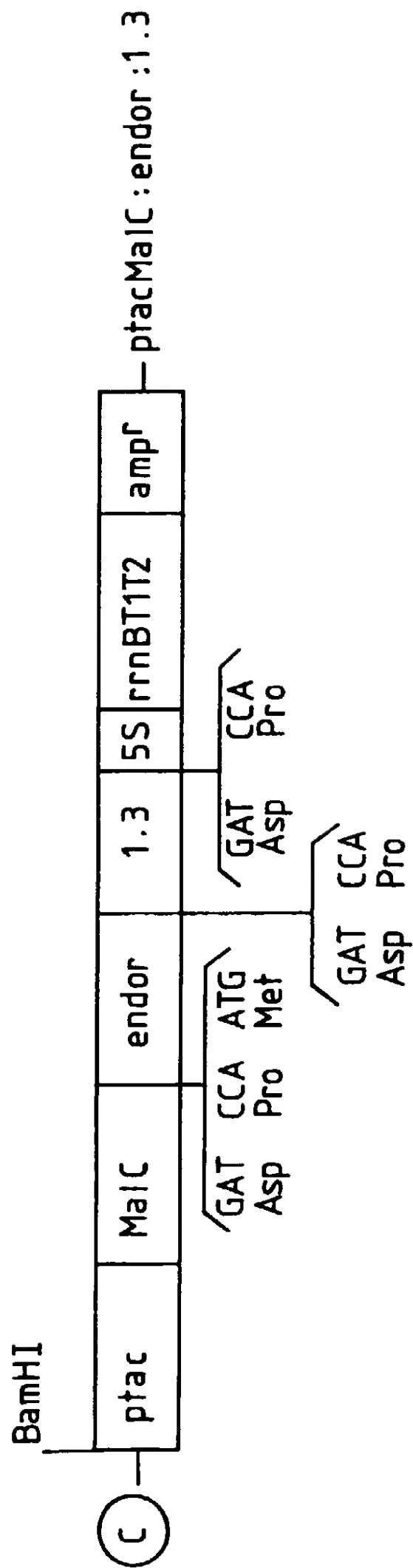
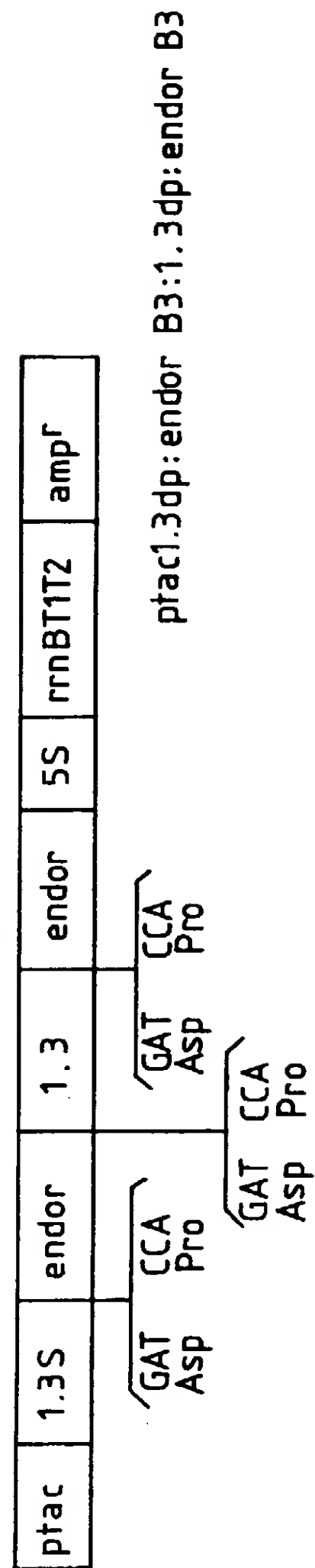


Fig.4





European Patent
Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 30 3067

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
D,X	WO-A-9 014 431 (BIOTECHNOLOGY RESEARCH AND DEVELOPMENT CORPORATION) 29 November 1990 * page 18, line 31 - line 34; claims 1-33; figure 2 *	1-25	C12N15/62 C07K3/18 C07K13/00 //C12N5/00
D,Y	J. BIOL. CHEM. vol. 265, no. 18, 25 June 1990, AM. SOC. MOL. BIOL., INC., US pages 10327 - 10333; J.E. CRONAN, JR: 'Biotination of proteins in vivo' * the whole document *	1-25	
Y	WO-A-8 909 393 (IGEN INCORPORATED) 5 October 1989 * the whole document *	1-25	
T	J. BIOL. CHEM. vol. 262, no. 24, 25 August 1987, AM. SOC. MOL. BIOL., INC., US pages 11813 - 11816; V.L. MUTIF AND D. SANDLS: 'Mutagenesis affecting the carboxyl terminus of the biotinyl subunit of transcarboxylase' * the whole document *	1-25	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)
			C12N C07K C12P
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 01 SEPTEMBER 1992	Examiner HORNIG H.
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons</p> <p>* : member of the same patent family, corresponding document</p>			

EPO FORM 1503 (01.82) (P0401)

-continued

1	5	10	15	
gac gtc gac aag tca cac gaa aac ccg atg ggc acc atc ctg ttc ggc				96
Asp Val Asp Lys Ser His Glu Asn Pro Met Gly Thr Ile Leu Phe Gly				
20		25	30	
ggc ggc acc ggc ggc gcg ccg gca ccg gca gca ggt ggc gca ggc gcc				144
Gly Gly Thr Gly Gly Ala Pro Ala Pro Ala Ala Gly Gly Ala Gly Ala				
35		40	45	
ggt aag gcc gga gag ggc gag att ccc gct ccg ctg gcc ggc acc gtc				192
Gly Lys Ala Gly Glu Gly Glu Ile Pro Ala Pro Leu Ala Gly Thr Val				
50		55	60	
tcc aag atc ctc gtg aag gag ggt gac acg gtc aag gct ggt cag acc				240
Ser Lys Ile Leu Val Lys Glu Gly Asp Thr Val Lys Ala Gly Gln Thr				
65	70	75	80	
gtg ctc gtt ctc gag gcc atg aag atg gag acc gag atc aac gct ccc				288
Val Leu Val Leu Glu Ala Met Lys Met Glu Thr Glu Ile Asn Ala Pro				
85		90	95	
acc gac ggc aag gtc gag aag gtc ctg gtc aag gag cgt gac gcg gtg				336
Thr Asp Gly Lys Val Glu Lys Val Leu Val Lys Glu Arg Asp Ala Val				
100		105	110	
cag ggc ggt cag ggt ctc atc aag atc ggg gat ctc gag ctc atc gaa				384
Gln Gly Gly Gln Gly Leu Ile Lys Ile Gly Asp Leu Glu Leu Ile Glu				
115		120	125	
ggt cgc gaa agc ttc agc tgg gat ccg gta ccg ata tca gat ctc cca				432
Gly Arg Glu Ser Phe Ser Trp Asp Pro Val Pro Ile Ser Asp Leu Pro				
130		135	140	
aaa aca cat aat act aaa gac aag ggt gct gaa gaa ctt gta aag tta				480
Lys Thr His Asn Thr Lys Asp Lys Gly Ala Glu Glu Leu Val Lys Leu				
145	150	155	160	
gct gaa tca gta gca ggc ttg cta aaa gta gcg caa gaa aca cta aat				528
Ala Glu Ser Val Ala Gly Leu Leu Lys Val Ala Gln Glu Thr Leu Asn				
165		170	175	
aat tca gtt aaa gaa ctt aca agt cct gtt gtg gca gaa agt cca aaa				576
Asn Ser Val Lys Glu Leu Thr Ser Pro Val Val Ala Glu Ser Pro Lys				
180		185	190	
aaa cct taacccgggg cgcccgcg				600
Lys Pro				

<210> SEQ ID NO 2

<211> LENGTH: 194

<212> TYPE: PRT

<213> ORGANISM: Borrelia burgdorferi

<400> SEQUENCE: 2

Met Lys Leu Lys Val Thr Val Asn Gly Thr Ala Tyr Asp Val Asp Val			
1	5	10	15
Asp Val Asp Lys Ser His Glu Asn Pro Met Gly Thr Ile Leu Phe Gly			
20		25	30
Gly Gly Thr Gly Gly Ala Pro Ala Pro Ala Ala Gly Gly Ala Gly Ala			
35		40	45
Gly Lys Ala Gly Glu Gly Glu Ile Pro Ala Pro Leu Ala Gly Thr Val			
50		55	60
Ser Lys Ile Leu Val Lys Glu Gly Asp Thr Val Lys Ala Gly Gln Thr			
65	70	75	80
Val Leu Val Leu Glu Ala Met Lys Met Glu Thr Glu Ile Asn Ala Pro			
85		90	95
Thr Asp Gly Lys Val Glu Lys Val Leu Val Lys Glu Arg Asp Ala Val			
100		105	110
Gln Gly Gly Gln Gly Leu Ile Lys Ile Gly Asp Leu Glu Leu Ile Glu			

-continued

115	120	125	
Gly Arg Glu Ser Phe Ser Trp Asp Pro Val Pro Ile Ser Asp Leu Pro			
130	135	140	
Lys Thr His Asn Thr Lys Asp Lys Gly Ala Glu Glu Leu Val Lys Leu			
145	150	155	160
Ala Glu Ser Val Ala Gly Leu Leu Lys Val Ala Gln Glu Thr Leu Asn			
165	170	175	
Asn Ser Val Lys Glu Leu Thr Ser Pro Val Val Ala Glu Ser Pro Lys			
180	185	190	

Lys Pro

<210> SEQ ID NO 3
 <211> LENGTH: 36
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR Primer

<400> SEQUENCE: 3

cgtggatcca tgaaaaagaa tacattaagt gcgata

36

<210> SEQ ID NO 4
 <211> LENGTH: 34
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR Primer

<400> SEQUENCE: 4

aattcccggg ttaaggtttt ttggacttt cgtc

34

What is claimed is:

1. An isolated, immunogenic polypeptide fragment of outer surface protein C (OspC) of *Borrelia burgdorferi* consisting essentially of an epitope of OspC having an amino acid sequence as shown in residues 145 to 194 of SEQ. ID. NO: 2.
2. An isolated polypeptide having an amino acid sequence as shown in SEQ. ID. NO: 2.
3. The isolated polypeptide of claim 2, further having an amino acid sequence as shown in residues 145 to 194 of SEQ ID NO: 2.
4. A pharmaceutical composition to vaccinate against and to treat borrelia infection in mammals, including humans, the composition comprising an amount of an isolated polypeptide consisting essentially of an amino acid sequence as shown in SEQ. ID. NO: 2, the amount being effective to vaccinate against or to treat borrelia infection in mammals.
5. The pharmaceutical composition of claim 4, further comprising a pharmaceutically suitable carrier.

6. The pharmaceutical composition of claim 4, wherein the isolated polypeptide consists essentially of an amino acid sequence as shown in residues 145 to 194 of SEQ. ID. NO: 2.

7. The pharmaceutical composition of claim 6, further comprising a pharmaceutically suitable carrier.

8. The pharmaceutical composition of claim 4, further comprising an adjuvant.

9. The pharmaceutical composition of claim 6, further comprising an adjuvant.

10. A kit for diagnosing borrelia infection in mammals, including humans, the kit comprising an isolated polypeptide having an amino acid sequence as shown in SEQ. ID. NO: 2, disposed in a suitable container therefor, and instructions for use of the kit.

11. The kit of claim 10, wherein the isolated polypeptide has an amino acid sequence as shown in residues 145 to 194 of SEQ. ID. NO: 2.

* * * * *